A METHODOLOGY FOR EVALUATING THE PERFORMANCE OF TOW-STEERED COMPOSITE TECHNOLOGY OVER A RANGE OF PLANFORM CONFIGURATIONS

A Dissertation
Presented to
The Academic Faculty

By

Coleby I. Friedland

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Aerospace Engineering

Georgia Institute of Technology

August 2021

A METHODOLOGY FOR EVALUATING THE PERFORMANCE OF TOW-STEERED COMPOSITE TECHNOLOGY OVER A RANGE OF PLANFORM CONFIGURATIONS

Approved by:

Professor Dimitri Mavris
School of Aerospace Engineering
Georgia Institute of Technology

Professor Graeme Kennedy
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Chung Lee
School of Aerospace Engineering
Georgia Institute of Technology

Professor Julian Rimoli
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Jesse Quinlan
Langley Research Center
NASA

Date Approved: June 3, 2021

ACKNOWLEDGEMENTS

I would first like to express my appreciation to my parents for their love, their foundational role in my life, and for the sacrifices they made that ultimately enabled my achievement. Similarly I would like to thank my brother Dawson, as well as my grandmothers, aunts, uncles, and cousins for their love, support, and inspiration over the years. I would especially like to thank my girlfriend Dr. Laura Weinstock for all the ways she has enriched my life, especially during this difficult and isolating past year and a half.

I would also like to thank all the members of my committee for sharing their expertise and endorsing my work. Their guidance and suggestions led to tangible improvements in my research and this dissertation. I am grateful to my advisor, Professor Dimitri Mavris, both for his support and for assembling the formidable team at ASDL. I greatly enjoyed working in the environment that he created. I am thankful to Dr. Chung Lee for regularly meeting with me and challenging me to express my ideas more clearly. I would like to thank Dr. Jesse Quinlan from NASA's Aeronautics Systems Analysis Branch for working with me and my lab and overseeing the project that inspired this work.

Additionally, I would like to express my appreciation to Dr. Jason Corman for his help with structural modeling, and to Dr. Jimmy Tai for his assistance with developing my ideas into my thesis work. Finally, I would like to thank all my friends, both from ASDL and elsewhere: Adam, Angela, Aroua, Brandon, Chelsea, Dan, Darshan, Dave, Gökçin, Jake, James, Jason, Justin, Leah, Natasha, Shai, Taylor, and others, for their encouragement and for the joy they have added to my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This thesis presents a methodology for evaluating the benefits provided by tow-steered composite technology over a range of wing planform configurations. Tow steering is the manufacture of composite structure with curved fiber paths. In contrast to conventional straight-tow manufacturing, this provides additional design freedom that can be taken advantage of to tailor structural load paths and, when applied to the structural wingbox, to grant passive aeroelastic load alleviation. These benefits can reduce structural weight and might enable more efficient, higher aspect ratio vehicle configurations.

In order to best take advantage of this technology, its system-level impact should be considered during the conceptual design stage. This impact can be implemented in a vehicle design tool via weight adjustment factors. However, the vehicle's configuration is still fluid at this point in the design, and the benefit from tow steering might depend on the planform. Therefore, the weight adjustment factor must be assessed across the range of potential planforms in order to achieve consistency between the system design and the technology performance.

As a new technology, there is a lack of historical data for the effects of tow steering and they must instead be calculated with physics-based simulation. This presents a challenge: engineering effort must be exerted developing and validating parametric physics-based models, and computational effort must be exerted running them at a variety of planform settings. Additionally, maximizing the technology benefit requires solving a high-dimensional constrained optimization problem to find the best structural design, including the tow paths, to use with each configuration. Approaches from literature solve this problem with gradient-based optimization algorithms but this imposes a requirement that the physics-based analysis efficiently calculates the needed derivatives, which might not be satisfied by available models or modeling frameworks.

In order to better leverage existing physics-based models and reduce the impact of

computational expense, the methodology pursues a surrogate modeling approach. Two obstacles to surrogate modeling were identified and became primary research focuses.

The first obstacle was dimensionality. Surrogate modeling techniques typically can be accurate for up to twenty to thirty input variables, but this design space has hundreds of input variables specifying the shape of the planform, the size of structural dimensions throughout the wingbox, and the tow paths themselves.

The active subspace method, which tries to find a subspace of important directions responsible for almost all of a function's variation, was investigated to determine if a tow-steered wingbox's behavior could be explained with a subspace small enough to permit surrogate modeling. However, this active subspace application was distinguished from previous ones by the need to identify a common subspace over all of the optimization problem's objective and constraint functions. An approach based on aggregating the objective and constraints using an augmented Lagrangian penalty function was developed and tested on a simple yet representative physics-based model.

The results showed that this approach was able to uncover an intuitive set of basis directions, but also that structural performance suffered from limiting the design to just this subspace. Therefore, a more conventional hierarchical decomposition technique was selected for the methodology. This technique enables surrogate modeling by only exposing a small number of high-level variables to the surrogate while implicitly accounting for the many local structural variables by setting their values in separate local optimization problems.

The second obstacle was gathering data for the surrogate model. Ideally, after sampling the physics-based model using a standard design of experiments the surrogate would be highly accurate. Unfortunately this was not the case, raising the question of how to choose further points to evaluate with the physics-based model and add to the surrogate.

Bayesian optimization methods answer this by identifying points that provide the most information about a problem's solution, and have been specialized for many types of prob-

lems. The goal of finding the best tow paths for a range of planforms is classified as a parameterized optimization problem, motivating the development of a Bayesian parameterized optimization method. Such a method was developed by modifying the knowledge gradient acquisition function used for optimization problems, and was tested on both a battery of artificial functions and the simple physics-based model.

The results on the artificial test functions showed that the Bayesian parameterized optimization method outperformed space-filling designs of experiments and tied the resulting surrogate's accuracy in regions of interest to the function's attributes. The results on the physics-based model demonstrated that the technique could efficiently find the family of optimal tow path designs across the planform space. As a result, the technique was incorporated into the methodology.

The findings from these efforts were synthesized to form the thesis methodology, which was then demonstrated in an example use case that applied the methodology to a wide-body transport aircraft. As part of following the methodology, a suitable environment for physics-based computational experimentation was created. This effort leveraged RADE, a software package for streamlining the construction of such environments that the author helped develop. After using Bayesian parameterized optimization to select the points to evaluate with the physics-based model, the surrogate model was shown to be accurate in the regions of the design space containing the likely minima.

An important result from the use case was that the weight reduction provided by tow steering depended nonlinearly on the planform, justifying the characterization of the technology benefit as a function of the planform. The weight reduction factors found in the use case were smaller than a comparable result from literature, suggesting that the technology benefit was limited by the coarse tow path parameterization that was needed to permit surrogate modeling. However, the results also suggest the existence of multiple local minima in some regions of the planform space that might lead a gradient-based optimization algorithm to settle in a slightly suboptimal local minima.

CHAPTER 1
INTRODUCTION

1.1 Motivation

Air transportation has a vital role in the world economy and is projected to keep growing in the long-term despite recent setbacks [1]. However, it causes adverse environmental impacts and as a result there is an ongoing effort to develop a new generation of aircraft that emit fewer pollutants, consume less fuel, and produce less noise. There are several projects that have been created in order to address these objectives, including NASA's Environmentally Responsible Aviation (ERA) project [2], NASA's Advanced Air Transport Technology (AATT) project [3], and the FAA's Continuous Lower Energy, Emissions, and Noise (CLEEN) program [4].

The AATT project is trying to attain specific performance metrics based on NASA's strategic goals, shown in table 1.1. These requirements are challenging to meet, especially when considering all of them simultaneously. For example, although it was expected that conventional aircraft concepts could be optimized to achieve individual N+2 level goals, it was also expected that these individually optimized concepts would have unacceptable performance towards the other goals [5]. In particular, the goals for noise and fuel burn conflict with each other.

If a feasible solution is not available within a conventional concept's design space, then one possible way to proceed is to infuse technologies. In that vein, tow-steered composite structures are being investigated for their potential to help achieve feasibility by reducing structural weight and enabling more efficient higher aspect ratio wings. Unlike conventional composite structures in which all the fibers are laid in straight lines, the fibers in tow-steered composite structures are arranged in curved paths with continuously varying orientations.

Table 1.1: NASA targeted improvements in subsonic transport system-level metrics [6]

Technology Benefits	**Technology Generations (Technology Readiness Level = 5–6)**		
	Near term 2015–2025	**Mid term 2025–2035**	**Far term Beyond 2035**
Noise (cumulative below Stage 4)	22–32 dB	32–42 dB	42–52 dB
LTO NO_x Emissions (below CAEP 6)	70–75 %	80 %	>80 %
Cruise NO_x Emissions (relative to 2005 best in class)	65–70 %	80 %	>80 %
Aircraft Fuel/Energy Consumption (relative to 2005 best in class)	40–50 %	50–60 %	60–80 %

With careful design of the tow paths, this extra freedom can be utilized to provide the benefits of passive aeroelastic load alleviation and improved resistance to load.

The direct effect of these benefits is to reduce the structural weight, which strongly influences the target metrics despite not being one itself. Therefore in order to determine how much adding tow-steered composites to a concept improves its feasibility, the structural effect must be propagated to the system level and considered alongside all the other major disciplines. This is a challenge because traditional conceptual design methods estimate the structural weight with statistically derived equations based on historical data, but there is no historical data for aircraft with tow-steered structures. As a result, physics-based analysis must be performed earlier in the design process to supply this missing information.

Physics-based analysis takes considerably more effort: all the details of fully realized aircraft that would be implicitly accounted for by a statistical equation must instead be explicitly modeled or abstracted away. While it might take a computer fractions of a second to calculate the weight with a statistical equation, physics-based analysis could take minutes

or hours. Additionally, the technology benefit should be characterized as a function of the planform for a couple of reasons. First, the final planform selection is not know because the gross design is still fluid during the conceptual phase. Second, an interaction between the wing planform and the wing weight reduction provided by tow steering is plausible. If the technology benefit is not characterized as a function of the planform, there is a risk that the vehicle design and the technology benefit will be inconsistent. Therefore, the physics-based model must support parametric geometry variation so that the weight reduction from tow steering can be evaluated over a range of planform settings.

The need to assess the impact of tow steering early in the design process and the challenges of developing an appropriate physics-based model and evaluating it over the range of potential planforms motivates the development of a systematic way to carry out these activities. This provides the motivation for this work, stated in the research objective:

RESEARCH OBJECTIVE

Develop a methodology to evaluate the wingbox structural weight reduction benefit provided by tow-steered composite technology across a range of planform configurations

1.2 Document Organization

Chapter 2 covers high-level background material relevant to the development of the desired methodology, including linking disciplinary technology effects to vehicle-level metrics early in design, structural technology performance evaluation, and tow-steered composite technology itself. A surrogate modeling approach is selected for the methodology, but a couple of gaps are identified and become the subject of further research in chapters 3 and 4.

Chapter 3 deals with the high dimensionality of the tow-steered composite design problem, which is an obstacle to surrogate modeling because surrogate models are typically able to cope with twenty to thirty variables at most. In response a dimensionality reduction method based on active subspaces but adapted for use on constrained optimization problems is developed. In order to assess the method, a simple physics-based experimental testbed is

created. Experimentation shows that the method is working correctly but is not effective on the target problem, and therefore a more conventional hierarchical decomposition approach is selected to manage the number of dimensions the surrogate acts on.

Chapter 4 addresses the selection of points for building the surrogate model. The behavior of the tow-steered wingbox is found to be too complicated to construct an accurate surrogate after an initial stage of data collection following a standard design of experiments. This creates a need to strategically select new points to add to the model. Bayesian optimization methods are found to provide a promising framework for selecting new points. However, because of the need to explore rather than optimize the planform settings, a specialized version for parameterized optimization is developed. Experimentation on artificial test functions and the previously constructed simple physics-based testbed suggests that the method is effective and should be included in the methodology.

Chapter 5 reviews the research formulation and experimental results before describing the thesis methodology. Chapter 6 presents an example use case that follows the methodology to calculate the weight reduction benefit from applying tow steering to a wide-body commercial transport aircraft. As part of the use case, a suitable environment for physics-based computational experiments is constructed. Afterwards, the results obtained by following the methodology are discussed. Finally, chapter 7 presents the conclusions and offers some ideas for future work.

CHAPTER 2
BACKGROUND

This chapter presents material that provides background on the problem of interest and can be used as a starting point for developing the thesis methodology, and then identifies a couple of gaps to be addressed by the work in subsequent chapters. First, it covers the connection between technology evaluation and the conceptual phase of aircraft design. Then, it reviews early design methods for structural weight estimation and structural technology performance estimation. Next, it examines the tow steering technology that is the focus of this work and surveys the literature on tow-steered composite wingbox design. Finally, it identifies challenges that this technology poses to the existing methods for structural technology performance estimation.

2.1 Technology Evaluation in Conceptual Design

Before diving into benchmark methods that can serve as a foundation for the thesis methodology, it makes sense to first consider where it will fit in the larger picture of aircraft design. This will expand on the motivation presented in the previous chapter and provide context for what the state of knowledge about a vehicle would be when the methodology is used.

To start, the methodology is intended for use in the conceptual design phase because that is the phase in which technologies are selected, and consequently there is a need to evaluate technologies within that phase. The conceptual phase [7] is the first phase of the aircraft design process, and usually begins with either a set of requirements from a customer or a guess as to what those requirements will be, though it can also start with a compelling idea for a vehicle. For this work, conceptual design starts with the ERA project's environmental impact requirements described in section 1.1. The critical question during the conceptual phase is whether any affordable aircraft can be built to meet the requirements.

The requirements drive an iterative evolution of the aircraft, characterized by a fluid process in which the design is constantly updated as trade studies are performed and new information is learned. Ultimately it arrives at a gross configuration arrangement that includes wing and tail planform geometry, fuselage shape, and the locations of payload and key subsystems, as well as the set of technologies that will be used. Making all these decisions requires evaluating a large number of alternative configurations in order to determine the general characteristics, size, and shape that provide the best overall vehicle performance.

One of the most important steps in the conceptual design cycle is sizing, which consists of translating a candidate vehicle's configuration data and mission requirements into size and system performance estimates. This calculation requires information about major aircraft subsystems such as aerodynamics, propulsion, and structures. Computer-aided tools known as sizing programs are used to automate this process, and in order to fit the nature of the conceptual phase they must support quick development, parametric capability, and rapid iteration. Therefore, subsystem behaviors are typically represented with simple equations based on empirical data or elementary analytical models [8]. Some examples of sizing programs are the Flight Optimization System (FLOPS) [9] and the more comprehensive Environmental Design Space [10], of which FLOPS is a part.

These tools were used to analyze conventional aircraft concepts with state-of-the-art technology and found that they would be unable to meet the ERA project's aggressive requirements for reduced environmental impact [5]. If there is no feasible solution in the conventional design space, then some ways to proceed are to 1) relax the constraints, 2) expand the design space, 3) select a different concept space, and 4) infuse new or alternative technologies [11]. Relaxing the constraints is obviously undesirable and might be out of the designer's control; this work focuses on the choice to both infuse tow-steered composite technology and expand the design space, since tow steering is thought to be an enabler for higher aspect ratio wings.

However, these options create difficulties by moving outside the existing knowledge base. The subsystem models used in conceptual design tools rely on historical data, assume that disciplinary requirements are satisfied, and assume that interdisciplinary interactions have been resolved. These simple models cannot reflect the use of a technology like tow steering, and might be invalid if their assumptions are violated or if the candidate design is dissimilar from historical designs. Additionally, trying to satisfy aggressive requirements causes an increased likelihood that the trends reported by the models push the design outside of their valid ranges. Because of these issues with the simpler models, more detailed physics-based analyses must be moved up earlier in the design process to improve reliability when advanced technologies or configurations are being considered.

Tow steering is not a panacea and will need to be evaluated alongside other technologies, possibly including other structural technologies, in order to find technology combinations with the best chance of meeting the requirements [12]. Since the technologies under consideration are inevitably immature (otherwise they would already be a part of the state of the art) there is significant uncertainty and risk associated with them, and there is a need to decide which technologies should be allocated resources to improve their maturity. It is nontrivial to find the best set of technologies: for a portfolio of n mutually compatible candidate technologies there are 2^n possible combinations (each can either be "on" or "off"). This exponentially large number of technology combinations combines with the configuration design space to create an enormous overall search space.

Looking through all of the exponentially many alternatives is intractable. The technology identification, evaluation, and selection (TIES) methodology [13] was developed in order to address the difficulty of searching through this enormous design space to select a technology package and configuration for an aircraft. Taking TIES as a representative approach, it consists of a set of steps that is helpful for describing how the thesis methodology will fit into this larger problem. The steps of TIES are: 1) define the problem, 2) define the concept space, 3) modeling and simulation, 4) investigate design space, 5) evaluate system feasibility,

6) identify technologies, 7) evaluate technologies, and 8) select technologies.

After the problem and concept space have been defined in steps 1 and 2, steps 3–5 model the vehicle at the system level to assess what fraction of the design space is feasible and how much improvement is needed. Then in step 6, after the set of candidate technologies has been identified: a) the technologies are checked for compatibility, b) the system impact of each technology are quantified, and c) the maturity of each technology is determined. In step 7 the technology impacts are propagated to determine the performance of candidate vehicles using the technologies, and in step 8 a set of technologies is selected based in part on how well they help to meet the requirements.

It is within step 6b, when the system impacts of each technology are quantified, that the thesis methodology is needed. Because the subsystem models used in conceptual design tools are usually unable to directly account for a technology, their outputs are modified with calibration or "k" factors instead. For instance, a propulsive technology could be represented as a percent reduction in specific fuel consumption or an aerodynamic technology could be represented as a percent reduction in cruise drag relative to the no-technology condition. Adverse effects and cost changes associated with a technology should also be accounted for.

Structural technologies like tow steering improve an aircraft by reducing its structural component weights. A concrete example of a suitable calibration factor to implement this effect is FLOPS's FRWI parameter, which adjusts the output of its wing weight model. If a more detailed breakdown is available then the parameters FRWI1, FRWI2, and FRWI3 can be used to calibrate the wing's bending (skin), shear (spar and rib), and other weight groups, respectively. At any rate, an appropriate set of structural weight calibration factors for tow-steered composites needs to be developed so that conceptual design tools can evaluate the performance of candidate aircraft using the technology.

The accuracy of an aircraft's projected performance and the suitability of a selected technology combination will naturally depend on the accuracy of the technology calibration factors. If there is significant uncertainty associated with the calibration factors, it should be

accounted for by modeling them as random variables and propagating their distributions to obtain probability distributions for system-level performance. Uncertainty in system-level performance can then be used as one of the criteria for selecting the vehicle's technology set and configuration. The calibration factors can be quantified by expert opinion, literature review, or physics-based modeling. If reliable technology performance estimates are not already available for a particular concept then physics-based modeling would be expected to provide the best estimates with the least uncertainty. However, it also requires the most resources.

As a final note, a potentially important aspect of a calibration factor's accuracy is its dependence on the configuration. The simplest approach is to assume that the technology impact is independent of the configuration and only evaluate it at a single baseline configuration. However, if this assumption is wrong, then the technology impact factors will be inaccurate for off-baseline configurations and cause the system-level performance estimates for those configurations to be incorrect. Since part of the benefit of tow steering comes from aeroelastic tailoring, there is good reason to think that its technology impact will depend on the size and shape of the wing, and that its calibration factor should be quantified as a function of the relevant configuration variables.

2.2 Early Design Wing Structural Weight Estimation

As shown in the previous section, there is a need to develop a calibration factor or set of calibration factors that represent the structural weight reduction provided by tow-steered composites during the conceptual design phase. This raises the questions of how structural weight is traditionally estimated early in the aircraft design process and why these methods are inadequate for this purpose.

The wing is often a point of special emphasis for weight estimation because it makes up a large portion of the total structural weight and has important interactions with other disciplines. The reason for this is that the wing receives the primary aerodynamic forces

and transmits them to the reacting weight forces. If the engines are mounted on the wing, then it also receives their propulsive force and weight. Finally, the wing is also usually used to store the majority of the fuel. As a result, at least for conventional tube-and-wing concepts, it is sensible to spend more effort and use more sophisticated methods for wing weight estimation than for other structural components.

There are a great variety of methods for estimating the wing's structural weight early in design, and authors surveying the field have developed their own classification schemes [14, 15, 16, 17]. The methods can roughly be arranged on a spectrum by complexity. The simplest methods are purely empirical: the most basic are based only on weight fractions of similar aircraft, and others expand on this by regressing over significant parameters such as design speed, load factor, and gross wing geometry. The next set of methods use elementary stress analysis to approximate the amount of structural material needed to resist the loads. The most complicated weight estimation methods use more detailed physics-based analyses such as classical plate theory or finite element analysis. Since these methods are applied early in design when many of the vehicle's details are still unknown and there is a need for rapid iteration, early-phase physics-based methods focus on the most important features/primary structure and use empirically derived adjustments to improve the accuracy of their predictions. In order to select a weight estimation technique one must consider how appropriate they are for a given concept, how far the design has progressed, and what trade-off between time and accuracy is desired.

Empirical Methods

Empirical methods regress weight data from a group of similar aircraft in order to develop a weight estimation equation for that class. The predictive factors are typically linked to the weight using a power law, and the power law's coefficient and exponents are fit using a procedure like least squares. Examples of these equations have been given by Murphy [17], Niu [18] and Raymer [19]. Murphy's formula, meant to be used for high-speed aircraft (like

fighter jets) with aluminum construction, is shown in equation (2.1). It is a simpler example of this kind of equation, as the parameters are grouped into a single predictive factor α before applying a power law.

$$WT = 110\alpha^{0.77} \times 10^{-6} \text{ lb} \tag{2.1}$$

$$\alpha = \frac{WTO\ XLF\ STSPAN\ SWING}{TROOT}$$

WT	wing weight (lb)
WTO	gross weight (lb)
XLF	ultimate load factor
$STSPAN$	structural span (ft)
$SWING$	gross wing area (ft^2)
$TROOT$	theoretical root thickness (ft)

This equation epitomizes the advantages of empirical weight regression methods: they can be evaluated instantly, they capture the relationship between the weight and important configuration variables or requirements, and they are very accurate when applied to similar aircraft concepts. This accuracy can be achieved with such little effort because empirical equations are in a sense reusing the tremendous effort that has already gone into the design of the fully realized aircraft in the regression data set. However, this kind of approach can only be used when there are a sufficient number of similar aircraft to build a valid regression model. It obviously cannot be used to forecast the impact of an advanced structural technology like tow-steered composites. A regression equation could serve as a basic structural weight model that technology calibration factors are applied to, but that still requires developing the calibration factors using a method driven by the relevant physics.

Analytical Methods

Analytical wing weight estimation methods are based on physics, but with such a low level of detail that they can often be expressed through closed form expressions. These methods start with an external force distribution consisting of an assumed lift distribution and possibly the weights of masses such as the engines and fuel, and scales the external force distribution by the takeoff gross weight and maximum load factor. The wing structure is modeled as a beam in order to derive cross-sectional shear forces and bending moments based solely on the external forces without needing to account for material stiffness, load paths, deformation, or aeroelastic interactions. The cross-sectional loads are then used to determine the stresses at each point along the span. By assuming that structural failure is caused by exceeding the material's strength, the required amount of structural material and the resulting weight of the primary structure can be determined.

An early example of an analytical method was developed by Micks [20], a more recent one was provided by Elham [15], and yet another instance of this approach is used for the wing weight calculations in the conceptual design tool FLOPS [21]. The formula for a relatively simple semi-analytical method developed by Torenbeek [22] is shown in equation (2.2).

$$W_w = 0.000912 k_{no} k_\lambda k_e k_{ue} [k_b n_w (W_{des} - 0.8 W_w)]^{0.55} \frac{b^{1.675}}{{\tau_r}^{0.45} (\cos \Lambda_{1/2})^{1.325}} + W_{hld} \quad (2.2)$$

W_w wing weight (lb)

k_{no} non-optimum correction factor

k_λ taper correction factor

k_e wing-mounted engine correction factor

k_{ue} wing-mounted undercarriage correction factor

k_b strut-braced wing correction factor

n_w ultimate load factor

W_{des} aircraft design weight (lb)

b wing span (ft)

τ_r root thickness-to-chord ratio

$\Lambda_{1/2}$ mid-chord sweep angle

W_{hld} high lift device weight (lb)

Since these methods use a simplified representation of the relevant physics, they should be calibrated using historical data. The non-optimum correction factor k_{no} is included in Torenbeek's equation for this purpose. This allows analytical and semi-analytical methods to be very accurate for existing designs just like empirical methods, but with the advantages that they can be more reliably extrapolated and that they can represent the effects of more configuration variables on the weight. Although they take more work to implement than empirical methods, once that is done their computations are easy enough that they can also be evaluated instantly.

Despite their advantages, these methods do not represent the physics with enough detail to account for tow steering's most important effects. Because they assume a lift distribution and neglect aeroelastic interactions, they are unable to show the benefit from passive aeroelastic load alleviation. Since internal forces are modeled at a cross-sectional level they are unable to depict local load path improvements.

Physics-Based Methods

Physics-based methods model the structure's geometry and behavior with comparatively more detail than the elementary analytical methods. Methods in this category have been based on either classical plate theory or the finite element method [23]. Unlike analytical methods, in which the internal loads only depend on the external forces, in these methods the internal loads also depend on the structure's stiffness properties. The external forces

themselves might depend on aeroelastic interactions. Physics-based methods also might consider more causes of structural failure than just exceeding the material's strength. Because of these differences, there is no longer a direct path from the external forces to the amount of structural material. As a result, physics-based methods must use a process such as an optimization algorithm to iteratively converge on the best feasible design while keeping the internal forces consistent with that design. The structural weight estimate is then the weight of the converged design, plus any calibration adjustments used to account for the model's limitations.

The finite element method allows for a lot of modeling flexibility, such as choosing which parts of the structure are explicitly represented and what types of elements are used. Not all are suitable at this stage in aircraft design. Giles [8] used the term "design oriented analysis" to refer to physics-based analysis that, while more detailed than traditional empirical and analytical methods, is still fast enough to use during conceptual design. He describes key characteristics that can be used to guide such analysis as a) adequate accuracy, b) efficient computation, c) the capability to trade accuracy for speed, d) minimal time for model preparation and modification, e) capability to generate sensitivity derivatives, and f) ease of coupling with other codes.

Along these lines, Giles created a method that represents the wing using several higher-order plate segments [24]. Finite element approaches in this category tend to be based on one-dimensional beam elements. Elham [25] created an aeroelastic method that represents the wing structure using beam elements coupled to an aerodynamic model. Bindolino [26] created a multilevel method that calculates cross-sectional forces with a global aeroelastic beam element model, and then transfers the cross sectional forces to more detailed local cross section analysis. Although used less commonly, models based on two-dimensional shell elements have an important advantage for tow-steered composites in that the variation in fiber alignment can be directly represented through the stiffness properties along both in-plane axes of the shell elements.

The complexity of models has increased over time, in part due to increases in computer power. For instance, it has been observed [27] that across decades the time it takes to evaluate structural models has stayed at a relatively constant six to eight hours, long enough to leave running overnight and check the results the next morning. Improvements in computational resources have been channeled into increasing fidelity and complexity rather than decreasing execution time.

2.3 Structural Technology Performance Estimation

As shown in the previous sections, there is a need to account for the use of tow-steered composite technology during the conceptual design phase, and this leads to the problem of using physics-based analysis to quantify a structural technology's impacts. This problem was considered by Corman in his dissertation [28], in which he described a general framework for linking multiple levels of analysis together. The highest level is conceptual design, in which the vehicle design space is explored in order to select a configuration and set of technologies, and determine the concept's feasibility. The lowest level of the framework is technology development and demonstration, in which the direct effects that a structural technology has on deformation, weight, strain, etc. are determined, possibly by physical experimentation. The middle level, technology performance estimation, links the low-level technology development and demonstration to the high-level conceptual design.

As the central component of the framework, it is the responsibility of technology performance estimation to receive information from both other components and transform it so it is usable by the opposite. Information about the vehicle configuration and mission profile from conceptual design must be combined with appropriate load cases [29] in order to define the local forces and boundary conditions that make up the technology's operating environment, for use in technology development and demonstration. Information about the structural properties from technology development and demonstration must be embodied in a structural sizing simulation in order to develop a set of weight reduction calibration factors

that can be used in conceptual design. Note that a technology's direct effect on weight cannot be used for the calibration factor because the structural design changes based on the technology's other effects on strength and stiffness.

Corman described a benchmark approach to technology performance estimation and then developed his own approach to address deficiencies he identified in the former. The benchmark approach came from work done to advance the Pultruded Rod Stitched Efficient Unitized Structure (PRSEUS) technology [30]. The goal of this process is to determine either the total structural weight reduction ΔW_S shown in equation (2.3a) or the structural weight reduction factor k_S shown in equation (2.3b), in which $W_{S,B}$ is the baseline structural weight and $W_{S,T}$ is the structural weight when using the technology. Since the chosen value will be used to modify a conceptual design tool's weight prediction, the baseline structural weight should correspond to that tool's weight prediction for consistency.

$$\Delta W_S = W_{S,B} - W_{S,T} \tag{2.3a}$$

$$k_S = \frac{W_{S,T}}{W_{S,B}} \tag{2.3b}$$

The technology performance estimation process presumes that a baseline vehicle configuration has been selected, then follows the steps: 1) generate the structural model, 2) size the structure, and 3) characterize performance. The structural model must be formulated such that it is able to account for the direct effects of the technology and should be at an appropriate level of detail based on the desired time/accuracy trade-off. In this application, it was found useful to model the stiffened panels making up the majority of the structure as shell elements with homogenized stiffness properties [31]. This allows the same finite element mesh to be able to represent structures with different stiffener designs. After the model has been generated, the structure is sized by minimizing the weight while honoring the structural failure and other constraints in all of the load cases. Since there are many

design variables controlling the panel thicknesses and stiffener dimensions throughout the structure, this optimization problem was split into separate component-wide optimization problems in which only local design variables and constraints were considered. However, because the design affects the load paths, the global finite element analysis and local sizing must be iterated to convergence. The result of sizing the structure is that the various technology-related properties for density, stiffness, strength, etc. are translated into a directly comparable overall weight value or set of group weight values.

Both the baseline and technology-infused structures should be sized using the same underlying model to make the weight comparison fair. However, more accurate weight data might be available for the baseline structure. In that case there will invariably be a disagreement between the data and the structural sizing model's predicted weight. This discrepancy is caused by limitations in the sizing model such as parts of the structure that are left out, failure conditions that are not checked, and missing weights (e.g., fasteners). When this happens, the sizing program's weights can be adjusted using non-optimum factors to ensure it gets the "correct" answer.

The final scaled weights of the baseline and technology-infused structures are then compared, possibly by group, to obtain the technology-enabled structural weight reduction. This value can then be used within the conceptual design tool to accurately account for use of the technology at the vehicle level.

The limitations Corman noted in the benchmark process were:

- The same structural topology is used for both the baseline and technology-infused structures, even though a different topology might better suit the technology-infused structure. This might be done for the purely pragmatic reason that automating the generation of structural finite element models is difficult, and it is much easier to reuse the same geometry and mesh. This could cause the technology to be unfairly penalized.

- The technology performance is reported as a point estimate, disregarding possible

dependence of the technology performance benefit on the vehicle configuration. This effect might be responsible for more technology performance uncertainty than the technology's immaturity. Like in the previous point this simplification could be motivated by the difficulty of automating the creation of finite element models for different outer mold line geometries, and the consequence could be the selection of a suboptimal conceptual design point. This concern is particularly important to the present work because of the possibility that tow steering's impact depends on the wing planform.

- The "true" non-optimum factors and strength knockdown factors for the technology-enabled case might differ from the baseline case. However, since there is no trusted data to derive them from like in the baseline case, they can only be appraised via expert opinion.

Clearly, one of the main sources of these limitations is the effort required to develop a suitable physics-based model. This creates a need for computational tools that can automate the production of these models, but building such a tool is itself a substantial task. The tool must manage an involved process composed of steps that on their own can be challenging. One of the first steps is generating the geometry. In order to study the effect of vehicle configuration and structural layout, the geometry must be parametrically generated based on a few high-level parameters. This can be difficult because as the planform and structural layout change the structural topology also changes, leading to different intersections between the structural components. There is also the possibly of pathological inconsistencies between the planform and structural layout definitions. After generating the geometry it must be meshed, which can also be difficult to automate. For example, Sensmeier, Stewart, and Samareh [32] developed a tool that used triangular finite elements for the wing skins because, although quadrilateral elements are preferred, mesh algorithms for quadrilateral elements failed too often. To complete the finite element model the tool must define the masses, initial element properties, boundary conditions, and load cases. The tool might need to interface

with aerodynamic software in order to obtain the aerodynamic portion of the loads. Finally, assuming that separate sizing software will be used to check if the internal loads cause structural failure, select stiffened panel designs, and homogenize panel stiffnesses, the tool will need to integrate and coordinate that software with the finite element solver.

The improved approach to technology performance estimation, named Structural Technology Evaluation for Experimental Design (STEED), explicitly accounts for the possibilities that the technology performance depends on the vehicle configuration or that it benefits from using a different structural topology. However, because it can take a lot of effort to develop the automated finite element models required to explore these dimensions, the method includes steps to consider the availability of existing models and to check whether running the model at multiple structural topologies or vehicle configurations is actually necessary. It proceeds in a bottom-up fashion from individual structural panels to the structural layout to the vehicle configuration, checking at each level whether use of the technology induces a different design. If not, then the simple single-point characterization of technology performance is used and the construction of a parameterized finite element model is skipped.

If a functional characterization is required, because each execution of the structural sizing program is expensive, STEED calls for the use of surrogate models to facilitate the comparison of the technology and baseline cases across different design points, and to visualize the relationships. Surrogate models are empirical models built from data generated by physics-based models. Ideally the surrogate captures the behavior of the original model, allowing for fast and accurate predictions across the range of the model inputs.

STEED was demonstrated by using it to obtain performance estimates of the PRSEUS technology, but could be applied to other structural technologies meeting a set of criteria. A suitable technology affects the weight indirectly by motivating changes to the structural design or even vehicle configuration, which necessitates the use of assembly- or vehicle-scale structural modeling to estimate the weight. The technology should also be adequately represented by homogenized shell properties. Tow steering meets these criteria, and as a

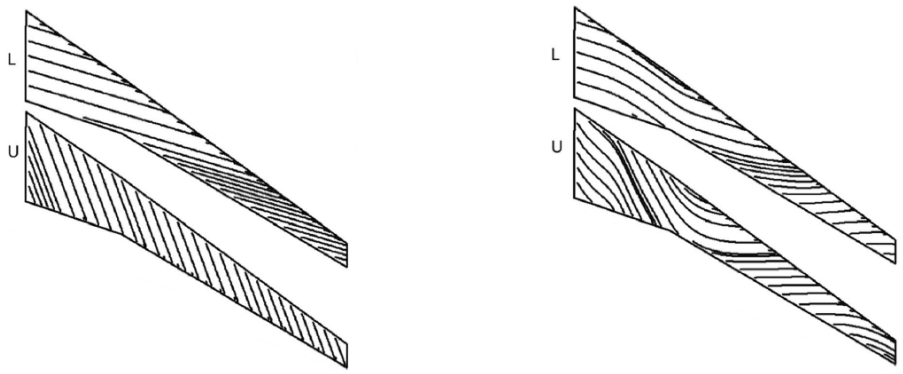

Figure 2.1: Straight and steered tow paths on the upper and lower surface of a wing [34]

result, STEED should provide a good starting point for the thesis methodology. However, the technology should also be considered in detail to determine if it poses any unique challenges.

2.4 Tow-Steered Composites

Tow-steered composite technology is being evaluated under the passive aeroelastic tailored wing category within the high aspect ratio optimal wing subproject of NASA's AATT project [33]. New manufacturing techniques allow the fibers in each layer of a composite laminate to be laid in curved paths with continuously varying orientations, unlike conventional composites in which all the fibers in a layer are straight and aligned in the same direction. Figure 2.1 shows a comparison of the tow paths in wing structures with and without tow steering. This technology provides the direct benefits of aeroelastic tailoring and local load path tailoring, which is expected to enable more efficient vehicle configurations with higher aspect ratio wings. Although this technology has not been used on any aircraft yet, a scaled-down tow-steered wingbox test article has been design, fabricated, and then load tested at NASA's Armstrong Flight Research Center, shown in figure 2.2.

Automated fiber placement (AFP) machines and software [36] are a key manufacturing technology that enable the fabrication of tow-steered composite structures, because of their ability to lay fibers in precisely programmed paths. There are some practical limitations, though. In locations where the tows converge/diverge, individual tows come together and

Figure 2.2: The passive aeroelastic tailored (PAT) wing bends under pressure from the highest loads applied during testing. **Credits: NASA / Ken Ulbrich** [35]

must either be dropped or overlap with each other, like in figure 2.3. AFP machines' capability to precisely cut and restart tows allows convergent tows to be dropped, but creates a gap in the ply. If the tows are overlapped instead, then the laminate will have an inconsistent thickness. In locations where the tows' curvature is too high, the bending of the tow into place by the AFP machine head can result in wrinkling. In order to avoid gap, overlap, and wrinkling phenomena, design limits can be placed on the convergence and curvature of tows [37]. Continuous tow shearing [38] is an alternate manufacturing technique for tow-steered composites that can reduce or eliminate these problems, although it is less mature than AFP.

One of the benefits of tow steering, aeroelastic tailoring [39], takes direct control of the coupling between bending and twisting of the wing. As an example, this can be used to design a wing that washes out under load, moving the aerodynamic forces inboard and thereby reducing the root bending moment. This would alleviate the aerodynamic maneuver

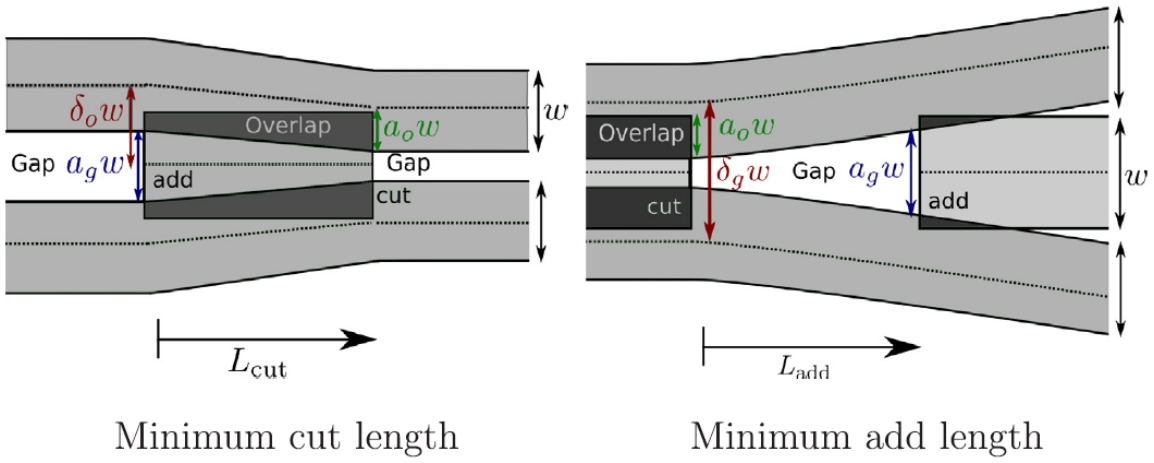

Figure 2.3: Gaps and overlaps due to tow convergence [37]

loads and increase the wing's resistance to static divergence failure. In tailoring the bend-twist coupling, there are subtle trade-offs between leading and trailing edge control authority, between load alleviation and lift effectiveness, and between static divergence and flutter. As a result, the best tow paths for a given wing are not obvious.

The other main benefit, local load path tailoring, is achieved by aligning the fibers with the local internal forces so they can react the load more efficiently. This allows strength- and buckling-based failures to be resisted with a smaller amount of structural material, and can also reduce stress concentrations near cutouts by redirecting the load more organically [40].

The tow paths need a mathematical definition to describe them, and in order to vary the design this definition must be parameterized. This can be accomplished with a tow orientation field that specifies the direction that the local tows are pointing for the entire surface of the structure. One of the simplest parameterizations for this field from literature [39] used only two parameters to define a spanwise linear variation in orientation angle from wing root to tip; the authors also experimented with adding a coordinate system tilt parameter to allow some chordwise variation but found that its effect was small. However, the majority of work on tow-steered wing skin design uses B-splines with a large number of coefficients for the orientation field. One of the most detailed parameterizations [41] used a B-spline

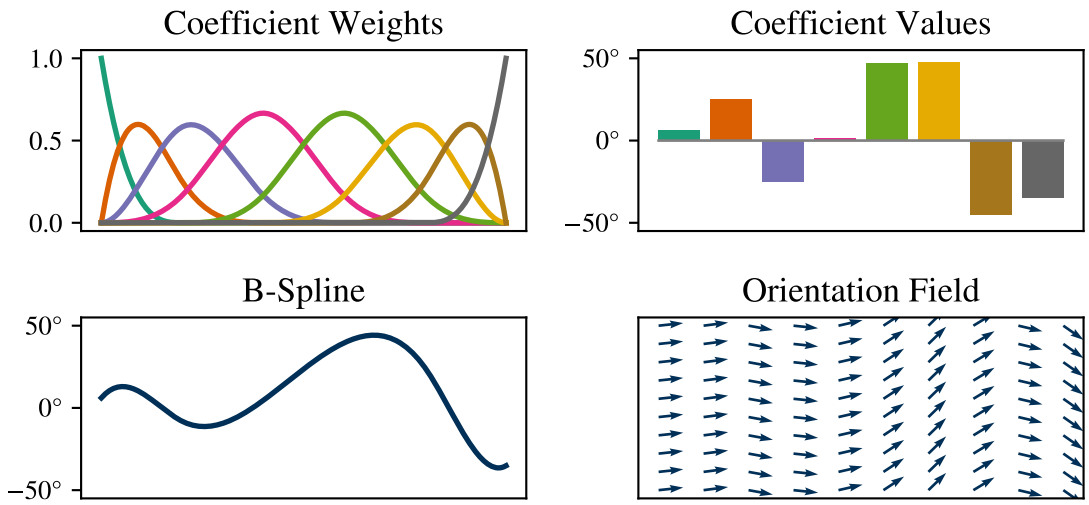

Figure 2.4: Example of tow orientation field parameterization. The upper-left plot shows the basis function associated with each coefficient of an eight-parameter cardinal cubic B-spline, and the upper-right shows randomly selected values for each coefficient. These coefficients are combined according to their basis functions to form the orientation function shown in the lower-left, which maps to the tow orientation field shown in the lower-right.

surface with 120 coefficients to specify the orientation field on each of the upper and lower wing surfaces. Other types of parameterized curves have been used in other applications of tow steering, but B-splines are a standard choice for wing skins. Figure 2.4 illustrates how a B-spline can be used to define a tow orientation field and thereby the tow paths. The tow orientation field also has a connection to the manufacturing constraints discussed a few paragraphs ago, in that the field's divergence is directly related to the convergence of the tows and the field's curl is directly related to the curvature of the tows. The field's divergence and curvature can therefore be used to express the manufacturing constraints.

In principle, every ply within a structure could have an independent tow path, but that would be impractical. First, since laminate wing structures have dozens of plies, the resulting design problem would be enormous. Second, it would be harder to certify the resulting structure. By regulation, aircraft structural designs are supported by physical tests in order to verify that their behavior is well understood and predictable [42]. These tests use a building-

block approach [43] comprised of tests at the coupon, element, detail, sub-component, and component levels. If the fiber orientations in each ply were allowed to vary independently, then there would be an intractably large number of laminate coupons that must be tested. To avoid this issue, a common approach is to define a single orientation field and a core laminate whose plies are set at fixed angles relative to the main orientation. In this manner, the same laminate can cover significant portions of the structure while at the same time the tow orientations vary continuously. For instance, the core laminate can be balanced, symmetric, and have a fixed ratio of 0°, ±45°, and 90° plies like in many ordinary composite laminates, but by continuously rotating this core laminate across the surface the structure becomes tow-steered.

2.5 Gap Identification

2.5.1 Gap 1: Dimensionality

Tow steering provides improvement by giving the designer extra freedom to take advantage of. As the dimensionality increases more benefit should be possible, albeit with diminishing returns. Methods from literature have maximized the technology benefit by using tow path parameterizations with dozens of variables, but using that many would be a challenge for the structural technology performance estimation methodology from section 2.3. That methodology fits a surrogate model over the vehicle configuration and structural topology variables, and handles the large number of design variables controlling stiffened panel dimensions throughout the structure by setting them in separate small local optimization problems. However, the many tow path variables should not be dealt with by including them in separate local problems because of their role in determining the wing's global aeroelastic response.

The remaining choices are then to either include the tow path variables in the surrogate model or to reformulate the structural design problem as a single large optimization of all of the stiffened panel dimensions and tow paths. If the first choice were selected, then the size

of the tow path parameterization would need to be capped because surrogate models are only able to cope with about twenty variables at most. This would be undesirable because it would limit the technology benefit.

If the second choice were selected, then gradient-based optimization would be the only practical way to solve the structural design problem, and the physics-based analysis would be required to provide efficiently calculated derivatives. This extra requirement is significant since one of the most important considerations for physics-based modeling early in design is development time. As an example, if a portfolio of technologies are being evaluated for an aircraft concept, it is possible that a parameterized structural model has already been developed to quantify the impact of another structural technology on that concept. It would take a relatively small amount of effort to modify this model so that its stiffness properties reflect the use of tow steering. However, if the existing model were not already formulated to efficiently calculate derivatives, then more extensive modifications would be needed to implement this capability. Even if an existing model were not available for reuse, having fewer requirements would help speed the development of the new model.

Lastly, there is a possibility of multiple locally optimal tow paths [44]. Although gradient-based optimization algorithms can efficiently find a local minimum, there is not a similarly efficient way to find the global minimum. Surrogate modeling could help by increasing the number of designs that can be evaluated for a fixed computational budget. For these reasons, it would be highly advantageous if a lower-dimensional tow path parameterization were able to realize similar technology benefits as a standard high-dimensional parameterization. This issue, which is addressed in chapter 3, is summarized in the following gap:

GAP 1

Existing approaches to tow-steered composite design need to use a high-dimensional tow path parameterization to maximize the technology benefit. However, this precludes surrogate modeling, requires gradient-enabled analysis, and makes it difficult to locate the global optimum.

2.5.2 Gap 2: Design of Experiments

Even if a new tow path parameterization that can capture the technology benefit with fewer design variables is found, unless it reduces that number to just a handful, there will still be a motivation to represent the tow paths with as many variables as possible in order to maximize the technology benefit. However, increasing the dimensionality makes surrogate modeling more difficult because more training data, obtained by evaluating the expensive physics-based model, would be needed to achieve acceptable accuracy. The exact amount of data required depends on the specific behavior of the original function, but a rule of thumb is that it becomes prohibitive at around twenty dimensions.

As a result, the dimensionality of the tow path parameterization controls a critical trade-off between technology benefit and computational cost. In order to maximize the efficiency of this trade-off, the data used to build the surrogate model must be collected as efficiently as possible. This can be accomplished by selecting a suitable design of experiments that reflects the expected system behavior and experimental goals, which for this problem have a unique feature that should be exploited: because the technology performance estimates are only based on the best, lightest weight tow path designs, the surrogate does not have to be accurate over all possible tow paths. In regions with poorly performing tow paths the surrogate only needs to be accurate enough to identify them as such. Therefore, by choosing a design of experiments that prioritizes more favorable regions of the design space, it should be possible to achieve a better technology benefit/computational cost trade-off. This is examined in chapter 4, where after not finding a design of experiments with the desired properties the following gap is addressed:

GAP 2

Because the trade-off between technology benefit and computational cost is significant, it must be made as efficient as possible. This creates the need to evaluate the expensive physics-based analysis at carefully selected points that reflect the system's behavior and the experimental goals.

CHAPTER 3
DIMENSIONALITY REDUCTION FOR STRUCTURAL DESIGN PROBLEMS

3.1 Background

The tow-steered composite design problem has parameters that continuously vary over the span of the wing such as panel thicknesses, stiffener dimensions, and the tow orientations themselves. Although the full set of possible designs could only be described with an infinite number of dimensions, for practical purposes a suitably large set of basis functions are chosen to provide design degrees of freedom. As the set of basis functions increases, the solutions of the finite dimensional problems should converge on the solution of the infinite dimensional problem. B-splines are a popular choice for the basis because their smoothness can be specified, the number of design variables easily be changed by adding or removing knots, and because each design variable's influence affects a localized region of the structure.

The downside of using B-splines or a similar piecewise polynomial basis is that it creates a high-dimensional optimization problem. Previous studies in tow-steered wingbox design [41] have had dozens or hundreds of variables. This presents a serious obstacle to global optimization due to the sheer volume of the search space. It also makes surrogate modeling and design space exploration infeasible, since they can cope with around twenty nontrivial design variables at most. As a result, gradient-based algorithms are all but required to solve these optimization problems, with no guarantees that the minimum found is the global minimum. It also requires that the physics-based analysis provide efficient gradient calculations. This would hinder the reuse of legacy models lacking this feature because they would need to be extensively modified or recreated. This is summarized in the following gap:

GAP 1

Existing approaches to tow-steered composite design need to use a high-dimensional tow path parameterization to maximize the technology benefit. However, this precludes surrogate modeling, requires gradient-enabled analysis, and makes it difficult to locate the global optimum.

3.1.1 Established Methods

High dimensionality is a feature of other structural design problems, so there is prior research on how to cope with it. One practical approach is fully stressed design [45], in which each component is sized to withstand the load applied to it with no extra margin. For statically determinate structures, this will immediately achieve the minimum weight design for the overall structure. In statically indeterminate structures, the load in each component depends on the design of the other components, so iteration is needed to converge the design and the load to a fixed point. Fully stressed design can be generalized a bit by considering multiple design variables and failure modes in each component. In this case, the minimum weight design with no negative failure margins is selected for each component in every iteration.

The advantage of a fully stressed design type approach is that the large optimization problem is decomposed into many small local optimization problems. This saves a lot of effort even when considering the iteration needed to converge the design and the load. This makes it possible to construct a surrogate model over some important global variables (e.g., planform variables like aspect ratio or structural layout variables like rib spacing) with individual panel thicknesses and stiffener dimensions set implicitly in the lower level problems.

The disadvantage of fully stressed design is that, unless all the degrees of freedom are constrained at the true optimum, the design it produces will be suboptimal. This might not be surprising if one notices that in its most basic form, fully stressed design does not even consider an objective function. Nevertheless it is still a common approach because it can

efficiently generate good, if not optimal, designs.

Fully stressed design could be applied to a tow-steered structure to enable surrogate modeling. However, only the panel and stiffener dimensions should be pushed down into the local subproblems. The tow orientations must be kept at the top level because some main benefits of tow steering, aeroelastic load alleviation and improved load path efficiency, act at a global rather than local level. Since the tow orientations will be kept at the top level, they must come out of the surrogate model's independent variable "budget," which limits how large the B-spline basis can be.

There are more sophisticated global-local methods that use multilevel [46] optimization to leverage the benefits of decomposition while preserving the true optimum. Like fully stressed design this produces many local subproblems but additionally has a global coordination subproblem to manage coupling between the subproblems. Unfortunately, the coordination problem must use design variables for each component (e.g., a weight budget and design load), so if there are many components then the dimensionality of the top level problem will by too high for surrogate modeling.

Another technique is to create a more specialized design basis. Pickett, Rubinstein, and Nelson [47] developed a reduced basis by combining the minimum gauge design with variable linking and quasi fully stressed designs, which are created by applying fully stressed design to individual load cases. Artificial load cases can be added to enrich the basis, though selecting the artificial load cases can be somewhat of an art. The authors found that optimizing in this reduced basis provided effectively the same result as optimizing the original design variables for a truss structure. If the optimum were also preserved in the case of tow-steered wingbox structures, and if the new design basis were small enough, then a technique like this would enable surrogate modeling with no loss in optimality.

Unfortunately, it is unclear how to adapt this approach for tow-steered composite design. As previously explained, (quasi) fully stressed design is not an appropriate way to choose the tow paths and there is no obvious way to link the tow path variables to other sets of variables.

One potential strategy would be to use quasi fully stressed design with straight tows to create a reduced basis of structural dimensions and combine that with a B-spline basis for the tow orientations. This would not capture interactions between the structural dimensions and tow orientations, and for surrogate modeling purposes the size of the structural dimension basis would need to be balanced against the size of the tow orientation basis. It is unclear whether this would perform better than creating a surrogate model over just the tow orientations while sizing the structural dimensions with fully stressed design.

A research question was developed to address the previously observed gap in light of these existing methods for dealing with high-dimensional structural design problems.

RESEARCH QUESTION 1

How should the tow-steered composite design problem be formulated and the tow paths be parameterized in order to facilitate surrogate modeling, while preserving as much of the technology's benefit as possible?

3.2 Method Development

Although the work by Pickett, Rubinstein, and Nelson is not directly transferable to tow-steered composite design, it does provide a compelling research direction. The high-dimensional B-splines used in previous tow-steered work are very flexible, but they are also very general. A specialized design basis might be able to capture important features of the problem more efficiently, leading to fewer design variables. This would allow surrogate modeling in the specialized basis without sacrificing much quality in the design, but raises the question of how to find a specialized design basis.

By applying dimensionality reduction techniques from outside the structures discipline, it might be possible to start from a high-dimensional B-spline basis and learn a more efficient specialized basis. Studying the reduced design basis might also provide insight about fundamental aspects of the problem that can be leveraged without having to repeat the dimensionality reduction procedure for every application.

Classical screening techniques look for the subset of independent variables that make significant contributions to the variability of the responses and eliminate those that do not. However, it is likely that almost all the B-spline coefficients for a tow-steered structure are significant, since the structural weight is the sum of all the component weights and every component must have positive failure margins. Even so, some groups of variables might have similar effects on the responses. That would create a structure in the problem that could be exploited by considering each significant group as a new independent variable.

3.2.1 Active Subspaces

The active subspace approach is based on finding an important subspace, or set of directions, rather than a subset of the independent variables [48]. Finding this subspace starts with the function of interest $f : \mathbb{R}^n \to \mathbb{R}$ which maps an n-dimensional input vector to a scalar output, and a weighting function $\rho : \mathbb{R}^n \to \mathbb{R}_+$ which specifies the region of interest. A simple choice for ρ is a uniform distribution over a scaled hypercube. The key quantity for determining the active subspace then is C, the uncentered covariance matrix of the gradient. Equation (3.1) shows how it is defined as the outer product of the function's gradient with itself, averaged over input space by the weighting function.

$$C = \int \nabla f(\boldsymbol{x}) \nabla f^{\mathrm{T}}(\boldsymbol{x}) \rho \, \mathrm{d}\boldsymbol{x} \tag{3.1}$$

Since this integral is taken over a large number of dimensions, it is recommended to estimate C by Monte Carlo integration using the weight function as a probability density [49]. The Monte Carlo estimate for C is given in equation (3.2), where each input sample $\boldsymbol{x}_j$ is drawn from the weight function. It is recommended [48] to use $N = \alpha k \log n$ samples, where n is the number of independent variables, k is a target number of dimensions for the low-dimensional model, and α is an oversampling factor. A recommended range for the

oversampling factor is two to ten.

$$C = \frac{1}{N}\sum_{j=1}^{N}\nabla f(\boldsymbol{x}_j)\nabla f^{\mathrm{T}}(\boldsymbol{x}_j) \tag{3.2}$$

The eigenvalues of C are then ordered from largest to smallest (all are positive because C is positive semidefinite) and examined for the presence of a large decrease between sequential eigenvalues. If there is, then the eigenvalues can be split at that point and there is an active subspace comprised of the eigenvectors associated with the large eigenvalues.

This works because the gradient information shows that the function varies far more in the directions corresponding to the large eigenvalues. The other directions can still have an effect but the loss in accuracy from ignoring them is relatively small. However, there is no guarantee that the size of the active subspace will be small enough for the desired purpose, or that there will be an active subspace at all. It completely depends on the function and the weighting. If there is no active subspace, then it still might make sense to focus on the largest eigenvalues, but the loss in accuracy from neglecting the remaining directions will be much higher. There is reason to think that the active subspace approach could be successfully applied to tow-steered composite design because it has been successfully applied to another high-dimensional nonlinear aerospace design problem, aerodynamic wing and airfoil shape design [50, 51].

3.2.2 Active Subspaces of Constrained Optimization Problems

The active subspace approach as described in literature cannot be directly applied to the tow-steered design problem because prior work has studied individual scalar functions but the tow-steered design problem is a constrained optimization problem with both an objective function and many constraint functions. A generic constrained optimization problem statement is shown in equation (3.3) with design variable vector $\boldsymbol{x} \in \mathbb{R}^n$, objective function f, and m inequality constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$ that must be satisfied at the solution. A more general problem statement could include equality constraints, but these

will be omitted for simplicity as there are none in the problem of interest.

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \leq 0 \quad i = 1, m \end{aligned} \tag{3.3}$$

A simple procedure would be to find individual active subspaces for the objective function and each constraint and then combine them into an overall active subspace, but this would likely be unsuccessful. The issue is that each failure constraint will most strongly depend on the local thickness B-spline coefficients, so the overall active subspace will end up looking very similar to the original parameterization and not provide any dimensionality reduction. To be successful, a method needs to capture how the design variables affect all of the problem's functions at the same time.

Another challenge is that the mesh and the structure's topology change along with the planform. As the span increases mesh elements and ribs spawn into existence and it is unclear how to continuously map their associate constraints. For these reasons, there needs to be a way to aggregate the objective and constraints into a single function for the overall problem.

The optimization literature has already developed a plethora of functions that combine the objective with a constraint-based penalty to produce a scalar value representing the entire problem. These are unsurprisingly called penalty functions [52], and are most apparent in sequential unconstrained minimization techniques [53], though they can also serve as merit functions to enforce the global convergence of other types of algorithms. The following subsections assess the suitability of these penalty functions for finding the active subspace.

Barrier Function

Constantine [54] suggests using a barrier function to aggregate the objective and constraints for the purposes of finding an active subspace. The log-barrier function shown in equation (3.4) is a representative example. This penalty function forces its minimum to be

feasible with respect to the constraints by approaching infinity as the design approaches the infeasibility boundary. When the penalty parameter is large, the barrier function's minimum is deep inside the feasible space. As the penalty parameter approaches zero, the barrier function's minimum approaches the solution of the original constrained optimization problem, though it also becomes very ill-conditioned.

$$\phi_{barrier}(\boldsymbol{x};\mu) = f(\boldsymbol{x}) - \mu \sum_i \log[-g_i(\boldsymbol{x})] \tag{3.4}$$

There are some issues with this penalty function. First, and a common weakness across all the penalty functions, is the fact that it depends on the penalty parameter μ to set the relative contributions of the objective and constraint functions. When penalty functions are used in the context of an iterative optimization algorithm, this parameter can be adaptively updated based on the algorithm's progress. In contrast, when a bunch of independent samples are taken, it is not clear how to find a good setting for this parameter. Additionally, it is sensitive to the relative scaling of the constraints, though that is a common challenge in optimization (e.g., if a constraint were multiplied by ten it would appear more important, even though the feasible space it allows would not change at all).

Looking at equation (3.4), it should be apparent that the logarithm terms cause the penalty function and its gradient to be undefined in the infeasible region. The only way to take the integral in equation (3.1) is to ensure that the weight function is zero in the infeasible region but that comes with challenges of its own. Since many of the constraints for tow-steered composite structures are nonlinear failure constraints, the weight function must be implicitly defined in terms of computationally expensive analysis. One way to do the Monte Carlo integration would be to reject samples with any failing constraints. Unfortunately, that would either lead to a large amount of rejected samples and wasted effort or an oversampling of more conservative regions of the design space.

Even if this problem were overcome, the extreme gradients near the feasibility boundary would still cause difficulties. Since the derivative of the logarithm diverges to negative

infinity near zero, in most cases the gradient of the penalty function will also approach infinity near the boundary. As a consequence, the integral in equation (3.1) might not converge, and if it does it might take a much larger number of samples to estimate accurately. And even if the integral converges and is accurately estimated, the much larger magnitude of the gradients near the boundary will tend to overemphasize that region when calculating the active subspace. It should be noted that these issues are purely due to the form of the penalty function and not due to the original problem itself.

Quadratic Penalty Function

A simple and popular penalty function is the quadratic penalty function $\phi_{quadratic}$ shown in equation (3.5). In contrast to the barrier function the quadratic penalty's minimum is far outside the feasible space (possibly unbounded) when the penalty parameter is small and approaches the solution of the original problem as the penalty parameter approaches infinity. Similarly to the barrier function, the function becomes ill-conditioned as the penalty parameter approaches its forcing value, and it is not clear how to set the penalty parameter for the purpose of uncovering the active subspace. Some clear advantages of the quadratic penalty function over the barrier function are that it is less nonlinear and does not restrict any of the domain.

$$\phi_{quadratic}(\boldsymbol{x};\mu) = f(\boldsymbol{x}) + \frac{\mu}{2}\sum_i \max[0, g_i(\boldsymbol{x})]^2 \tag{3.5}$$

Augmented Lagrangian

The augmented Lagrangian [55] shown in equation (3.6) is formed by adding a quadratic penalty term to the original optimization problem's Lagrangian, or equivalently by the multiplier-scaled constraint term from the Lagrangian to the quadratic penalty function. In this way it can be seen as an extension of the quadratic penalty function. In addition to the penalty parameter, it takes a vector of Lagrange multipliers $\boldsymbol{\lambda}$ as an argument. Its gradient,

which is what would be sampled to find the active subspace, is shown in equation (3.7).

$$\mathcal{L}_{\mathrm{A}}(\boldsymbol{x}, \boldsymbol{\lambda}; \mu) = f(\boldsymbol{x}) + \sum_i \left(\lambda_i \hat{g}_i(\boldsymbol{x}) + \frac{\mu}{2} \hat{g}_i^2(\boldsymbol{x}) \right) \tag{3.6}$$

$$\text{with} \quad \hat{g}_i(\boldsymbol{x}) = \max\left(g_i(\boldsymbol{x}), -\frac{\lambda_i}{\mu} \right)$$

$$\nabla_{\boldsymbol{x}} \mathcal{L}_{\mathrm{A}}(\boldsymbol{x}, \boldsymbol{\lambda}; \mu) = \nabla f(\boldsymbol{x}) + \sum_i \left(\big(\lambda_i + \mu \hat{g}_i(\boldsymbol{x})\big) \nabla \hat{g}_i(\boldsymbol{x}) \right) \tag{3.7}$$

$$\text{with} \quad \nabla \hat{g}_i = \begin{cases} \nabla g_i(\boldsymbol{x}) & \text{if } g_i(\boldsymbol{x}) > -\frac{\lambda_i}{\mu} \\ 0 & \text{otherwise} \end{cases}$$

The form of the augmented Lagrangian used here is somewhat nonstandard compared to its typical use within optimization algorithms: in that setting, each inequality constraint is converted to an equality constraint by adding a bound-constrained slack variable. But for the purpose of finding the active subspace, these slack variables would be a nuisance with no benefit, cluttering and confusing what the components of the gradient samples represent. The reformulation applied here eliminates the slack variables in a way that minimizes the standard augmented Lagrangian while satisfying the bound constraints.

The augmented Lagrangian is similar to the quadratic penalty function but has some advantages. It is less sensitive to the scaling of the constraints because the products of the multipliers and the constraints depend only the scale of the objective function and the inherent geometry of the problem. If a constraint were multiplied by ten then its associated multiplier would decrease by a factor of ten and their product would stay the same. Another advantage of the augmented Lagrangian is that the penalty parameter does not need to approach infinity for the minimum to match the solution of the original optimization

problem, it only needs to be large enough for the augmented Lagrangian to have positive definite curvature. If it already has positive definite curvature, then the penalty term is not needed. Because of this, the augmented Lagrangian is less sensitive to the choice of penalty parameter and it is less likely that a large value that causes ill-conditioning would be used.

The difficulty in using the augmented Lagrangian is in determining the Lagrange multipliers. These are not directly available for general nonlinear programming problems and must be iteratively calculated along with the solution. Within an active subspace method, the multipliers will need to be estimated based on the information obtained at each sample. It seems likely that the accuracy of the multiplier estimates would affect the success of the overall method. A procedure for doing so is described in subsection 3.2.3. This estimation procedure adds significant computational expense compared to the other penalty functions, which could be an important consideration depending on how expensive the samples are to obtain.

Review of Penalty Functions

Several penalty functions have been examined for their suitability for finding the active subspace of a constrained optimization problem. Their strengths and weaknesses are summarized in table 3.1. All of them have some sensitivity to the scaling of the constraints and the penalty parameter, so an engineer using this approach will need to be aware of this and be prepared to tune them for best performance. The augmented Lagrangian should be less sensitive to these concerns.

The barrier function creates severe ill-conditioning near the feasibility boundary and is not defined outside the feasible space. Since these issues make it difficult to use it is not considered further in this research. The augmented Lagrangian, which seems the most promising, and the simpler quadratic penalty are evaluated in section 3.4 to determine their effectiveness at reducing the dimensionality of the tow-steered composite design problem.

Table 3.1: Qualitative assessment of penalty functions for finding the active subspace of a constrained optimization problem

Criteria	**Logarithmic barrier**	**Quadratic penalty**	**Augmented Lagrangian**
domain restriction	✗	✓	✓
ill-conditioning	✗	—	✓
constraint scale sensitivity	✗	✗	—
penalty parameter sensitivity	✗	✗	—
additional computation	✓	✓	✗

3.2.3 Estimating the Lagrange Multipliers

Sampling the gradient in equation (3.6) requires the values of the constraint functions, the gradients of the objective and constraint functions, and the values of the Lagrange multipliers. The objective and constraint function values are straightforwardly obtained by running the analysis, but the true multipliers can only be obtained by iteratively solving the optimization problem. However, it would be prohibitively expensive if to fully converge a solution for each sample, so the multipliers must be estimated instead.

Although estimation of the multipliers was glossed over in the previous section, it is an important detail of the procedure. Many optimization algorithms need to estimate the multipliers and do so by solving a simplified subproblem. Equation (3.8) shows a subproblem formed from an assumed curvature matrix B, which will be elaborated on later, and a linearization of the original problem's objective and constraints centered at the sample. This forms a quadratic programming problem that can be solved using standard techniques [56]. The multipliers from the solution of the subproblem can then be used as

multiplier estimates for the original problem.

$$\begin{aligned} \min_{\Delta \boldsymbol{x}} \quad & \frac{1}{2}\Delta \boldsymbol{x}^{\mathrm{T}} B \Delta \boldsymbol{x} + \nabla f(\boldsymbol{x})^{\mathrm{T}} \Delta \boldsymbol{x} \\ \text{such that} \quad & g_i(\boldsymbol{x}) + \nabla g_i(\boldsymbol{x})^{\mathrm{T}} \Delta \boldsymbol{x} \leq 0 \quad i = 1, m \end{aligned} \tag{3.8}$$

The rest of this subsection will discuss some complications involved in solving this subproblem: determining the curvature matrix and handling the case that the subproblem is infeasible.

Curvature Matrix

Ideally, the curvature matrix used in equation (3.8) would be the Hessian of the Lagrangian (modified to ensure it is positive definite) at the sample point, but that would require the Hessians of the objective and constraint functions which are not typically available. Even if they were, it would be expensive and same effort used to evaluate the Hessians would likely be better spent if it were used to increase the number of samples instead.

In the absence of Hessian information, it might seem most appropriate to assume no curvature and to use the zero matrix. In that case, equation (3.8) becomes a linear programming problem. However, there are good reasons to assume positive curvature. First, at the solution of a nonlinear optimization problem, the curvature will be positive semidefinite in the nullspace of the active constraints. Second, the tow-steered composite design problem is known to be nonlinear, so it has curvature. In particular, stress- or strain-based failure constraints have positive curvature with respect to the thickness. Third, a consequence of assuming zero curvature is that the solution of the subproblem (3.8) will tend to an extreme corner of the design space. This would not represent the underlying problem very well: intuitively, the optimal structural thicknesses and the tow orientations should be somewhere in the middle of the design space. Additionally, at extreme values the linearized objective and constraints are less accurate. Assuming curvature provides a soft penalty that discourages extreme solutions.

Although there is a good case for assuming positive curvature, it less clear what the magnitude of the positive curvature should be. In the absence of more specific information, it is possible to set the curvature based on the scale of other problem quantities. The idea behind this is that, to arrive at the optimum from a randomly sampled point, the magnitude of the gradient needs to vanish over a step whose length is similar to the length of the design space.

The weight function used in the active subspace approach can be reused to derive a diagonal input scaling matrix $S_{\boldsymbol{x}}$ by taking the standard deviation of its marginal distribution in each dimension. A scale for the objective function s_f can then be set to the magnitude of a typical elements of the input-normalized gradient $s_f \sim \frac{1}{\sqrt{n}}\|S_{\boldsymbol{x}}\nabla f(\boldsymbol{x})\|$. The desired property can then be attained by assuming a curvature $B = s_f S_{\boldsymbol{x}}^{-2}$. Note that if the input scale is the identity matrix I and the objective function scale is 1, then the assumed curvature will be the identity matrix I. This is a rough heuristic that can be used without any additional information or insight, so it is likely that a problem-specific value will perform better.

Constraint Relaxation

Another difficulty that can arise is that the linearized constraints in subproblem (3.8) might allow such a small feasible space that the solution is forced into an extreme and unrealistic corner of the design space. There might even be no feasible space so that the subproblem has no solution. That does not necessarily mean that the original problem's solution is in an extreme corner or does not exist, but it does present an obstacle to estimating the Lagrange multipliers.

The remedy presented here is based on a constraint relaxation idea from Byrd, Schnabel, and Shultz [57]. The intuition behind it is that even if the linearized problem has no reasonable solution, it should still be possible to find an improvement that reduces the constraint violation and also potentially the objective.

First, a trust region with radius Δ is introduced, and the auxiliary subproblem shown in

equation (3.9) is solved to find the minimum violation of the linearized constraints possible within a reduced trust region. The solution of this subproblem gives a relaxation vector with elements $r_i = g_i(\boldsymbol{x}) + \nabla g_i(\boldsymbol{x})^{\mathrm{T}} \Delta \boldsymbol{x}^*$. If there is feasible space within the reduced trust region, then all the elements of the relaxation vector will be zero.

$$\begin{aligned} \min_{\Delta \boldsymbol{x}} \quad & \frac{1}{2} \sum_i \max\bigl(0, g_i(\boldsymbol{x}) + \nabla g_i(\boldsymbol{x})^{\mathrm{T}} \Delta \boldsymbol{x}\bigr)^2 \\ \text{such that} \quad & \|\Delta \boldsymbol{x}\| \leq 0.8\Delta \end{aligned} \tag{3.9}$$

The second step is to solve the relaxed linearized subproblem within the trust region, shown in equation (3.10). This subproblem will always have a solution, for instance, the solution from the previous auxiliary can be used as a feasible starting point because it is feasible with respect to the relaxed constraints and the trust region constraint.

$$\begin{aligned} \min_{\Delta \boldsymbol{x}} \quad & \frac{1}{2} \Delta \boldsymbol{x}^{\mathrm{T}} B \Delta \boldsymbol{x} + \nabla f(\boldsymbol{x})^{\mathrm{T}} \Delta \boldsymbol{x} \\ \text{such that} \quad & g_i(\boldsymbol{x}) + \nabla g_i(\boldsymbol{x})^{\mathrm{T}} \Delta \boldsymbol{x} \leq r_i \quad i = 1, m \\ & \|\Delta \boldsymbol{x}\| \leq \Delta \end{aligned} \tag{3.10}$$

The spherical trust region added by this constraint relaxation procedure changes the subproblem from a quadratic programming problem to a more difficult nonlinear programming problem. However, because it uses linearizations of the objective and constraints rather than the expensive analysis itself, this should be inconsequential. Finally, and similarly to the curvature issue, there is not an obvious setting for the trust region radius. Also similarly to the curvature, a good guess is that it should be the same order of magnitude as the design space.

3.2.4 Method Description

Because standard parameterizations used for tow-steered composite design create high-dimensional optimization problems, design space exploration is impossible and gradient-enabled analysis is required. The well-known fully stressed design technique could be used

to get around this but leads to a suboptimal result. Since the high dimensionality primarily comes from B-spline coefficients it is speculated that a more efficient parameterization, tailored to the behavior of a tow-steered composite wingbox, would enable similar quality solutions to be obtained with far fewer dimensions.

Based on the reasoning in section 3.2, a method has been developed based on the active subspace idea, but adapted for a constrained optimization problem. Although this method requires gradients, it is anticipated that by applying it to a simple gradient-enabled model useful basis designs and insight about independent variable interactions can be learned. That new information could then be leveraged to craft low-dimensional parameterizations for other analysis models, in particular more complicated ones that do not provide efficient gradient calculations.

The overall procedure is given by algorithm 1. Each gradient sample requires one evaluation of the objective and constraint functions and their gradients, as well as the solution of two nonlinear programs. If the analysis used to evaluate the objective and constraints is expensive (like for tow-steered composite structures), then it should dominate the overall cost and the extra effort to solve the nonlinear programs is negligible.

Algorithm 1 Finding the active subspace of a constrained optimization problem

Require: Constrained n-dimensional optimization problem with objective function $f(\boldsymbol{x})$ and constraint functions $\boldsymbol{g}(\boldsymbol{x})$, weight density over the input dimensions $\rho(\boldsymbol{x})$, target reduced number of dimensions k, oversampling factor α, curvature matrix B, and penalty parameter μ

1: Determine the number of samples from $N = \alpha k \log n$
2: **for** $j = 1, N$ **do**
3: Draw a random sample $\boldsymbol{x}$ from ρ
4: Evaluate $\nabla f(\boldsymbol{x}), \boldsymbol{g}(\boldsymbol{x}),$ and $\nabla \boldsymbol{g}(\boldsymbol{x})$
5: Estimate the Lagrange multipliers $\boldsymbol{\lambda}$ either from solving subproblem (3.8) or, if subproblem feasibility is an issue, subproblems (3.9) and (3.10)
6: Set gradient sample $\nabla \mathcal{L}_{\mathrm{A}j}$ using equation (3.7)
7: **end for**
8: Assemble gradient samples into uncentered covariance C using equation (3.2)
9: Take the eigendecomposition of C, and examine the sorted eigenvalues for the presence of a gap in the magnitude of subsequent eigenvalues
10: Decide how many eigenvectors to use in the lower-dimensional representation

The reasoning presented in this section is embodied in the following hypothesis statement, which is tested in the next section.

HYPOTHESIS 1
If the objective and constraint functions of the tow-steered composite design problem are aggregated by an augmented Lagrangian penalty function, and if the active subspace of this penalty function is found, then that active subspace can be used as a low-dimensional parameterization of the problem suitable for surrogate modeling without sacrificing much performance. This is because the penalty function's minimum is the solution of the original problem and because the active subspace approach is able to find a reduced set of directions that explain most of a function's behavior.

3.3 Physics-Based Testbed

At this point, with the proposed method's reasoning laid out and a hypothesis about the method formed, the next step is to test that hypothesis. In order to do so a suitable experimental testbed, which will take the form of a physics-based analysis program, must be created. Because the proposed active subspace method, and the larger thesis methodology, are being developed to address the specific features of tow-steered composite technology impact estimation, the testbed must capture all of the important physical behaviors that affect actual wingbox structures. At the same time it should be as simple as possible and have a minimal computational expense. This will allow for easy data generation and prevent unnecessary details from complicating analysis and interpretation of results.

3.3.1 Formulation

The externally applied forces come from the air flowing around the wing and can be understood using aerodynamics. These external forces cause internal forces and deformations in the structure that are described by solid mechanics. There is coupling between aerodynamics and solid mechanics, referred to as aeroelasticity, because while the airflow and

external forces depend on the deformation of the wing, at the same time the deformation of the wing depends on the external forces. Although aeroelasticity is sometimes ignored in physics-based wing weight estimation methods for simplicity, it is necessary here because aeroelastic tailoring is an important benefit from tow steering. Therefore, it must be accounted for when evaluating the technology benefit. Finally, structural failure theories are needed in order to determine whether the internal forces and deformations cause the structure to break. To create the physics-based model, these behaviors must be described mathematically, assembled together, and implemented in such a way that they can be solved numerically.

Aerodynamics

The aerodynamic forces applied to the wing are fundamentally governed by the conservation of mass and the conservation of momentum [58]. These laws are embodied mathematically through the continuity equation and a set of three momentum equations known as the Navier-Stokes equations.

The highest-fidelity approaches to aerodynamic analysis are computational fluid dynamics (CFD) methods, which approximately solve the governing equations on a three-dimensional grid. However, they are so computationally expensive that they are impractical for early design structural weight estimation. These are the only methods that can accurately calculate the drag force, but other methods are able to accurately calculate the lift with much less expense by making simplifying assumptions such as potential flow. Because the lift force on the wing is much larger than the drag force, it is more important to structural sizing and these other methods can provide aerodynamic forces with adequate accuracy for structural weight estimation.

One of the simplest approaches is to assume a spanwise lift distribution like in Schrenk's method [59]. In Schrenk's method, the lift distribution is the average of an elliptical distribution and the planform's cross sectional area distribution, possibly with further

adjustments to account for the wing's twist. This distribution is scaled in each load case based on the load factor and an assumed wing lift fraction to get the right total lift.

Vortex lattice methods [60, 61] are more sophisticated than assuming a distribution but still substantially faster than CFD. These methods idealize the wing (or other aerodynamic surface) as a flat surface and discretize it into a lattice of quadrilateral panels. Each panel has both a horseshoe vortex and a control point associated with it. The horseshoe vortices contribute to the overall flow field and are assumed to leave the wing in flat sheets, but their strengths are unknown. The vortex strengths are linked to the normal flow (perpendicular to the local mean camber line) at each control point via aerodynamic influence coefficients to create a system of linear equations. Solving this system of equations for the condition where the normal flow is zero at every control point enforces flow tangency boundary conditions and yields the vortex strengths defining the flow field. Because the overall flow is modeled as a superposition of vortices, it will automatically be a potential flow and satisfy the constant-density continuity equation. The pressure over the wing surface can then be calculated from Bernoulli's equation (inviscid incompressible momentum). In the high subsonic regime the Prandtl-Glauert correction is used to account for compressibility effects. Because vortex lattice methods have a favorable trade-off between accuracy and computational expense for structural weight estimation, the testbed will use a vortex lattice method to calculate the aerodynamic forces.

Inertial forces are another important source of loads. These come from the primary structure itself, wing-stored fuel, wing-mounted engines (which would also contribute a thrust force), secondary structure, and other subsystems located in the wing. These are easy to calculate since they only depend on the mass and the acceleration relative to gravity. For simplicity, the testbed only includes the mass of the primary structure.

Solid Mechanics

Solid mechanics theory [62] predicts how a structure deforms under external load and what internal loads result. Maneuver load cases are commonly modeled with static analysis which is based on an equilibrium state of the structure. Dynamic analysis is needed for gust load cases and to model phenomena like flutter and divergence. However, because of the expense, structural weight prediction methods for early design typically do not use dynamic analysis.

The governing equations for static structural deformation are the equilibrium equations, the constitutive law, and the strain compatibility equations. The equilibrium equations enforce a static state of the structure, the strain compatibility equations ensure that the strain field corresponds to a valid displacement field, and the constitutive law defines the relationship between stresses and strains. For early structural design, it is usually assumed that deformations are small and therefore that the strain-displacement and stress-strain relationships are linear. This assumption is reasonable because structures are typically designed to not experience large deformation under normal operating conditions.

For a cube shaped differential element of the structure, the elements of the stress tensor σ_{ij} represent the stress acting on the face normal to direction i in direction j. For the structure to be in static equilibrium, it must be in equilibrium both within its volume and on its surface. A differential element in the volume of the structure is in equilibrium when the difference in stress across it balances the body forces (such as inertial forces), as shown in equation (3.11a) where b_j is the component of the body force per volume in direction j. A differential element on the surface of the structure is in equilibrium when the external traction matches the stress on the face of the surface, as shown in equation (3.11b) where t_i is the external traction in direction i and n_j is the jth direction cosine of the surface normal. Additionally, there must be no net torques within the structure with the consequence that the

stress tensor is symmetric.

$$\sum_{i=1}^{3} \frac{\partial \sigma_{ij}}{\partial x_i} + b_j = 0 \quad \text{for} \quad j = 1, 2, 3 \tag{3.11a}$$

$$t_i = \sum_{j=1}^{3} \sigma_{ij} n_j \quad \text{for} \quad i = 1, 2, 3 \tag{3.11b}$$

Under linear assumptions, the elements of a solid's strain tensor are given by equation (3.12), in which $\boldsymbol{u}(\boldsymbol{x})$ is the displacement field.

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{3.12}$$

A structure's deformation is uniquely defined by the three component displacement field and it is seen that this displacement field induces a symmetric strain tensor field with six independent components. It is therefore straightforward to calculate the strain field from the displacement field, but the inverse is not. The derivation of six strain components from only three displacement components implies that only three of the strain components are independent. The remaining components are constrained by strain compatibility equations, of which two ultimately equivalent forms are shown in equation (3.13).

$$\frac{\partial^2 \epsilon_{ij}}{\partial x_i \partial x_j} = \frac{1}{2}\left(\frac{\partial^2 \epsilon_{ii}}{\partial x_j^2} + \frac{\partial^2 \epsilon_{jj}}{\partial x_i^2}\right) \quad i \neq j \tag{3.13a}$$

$$\frac{\partial^2 \epsilon_{ii}}{\partial x_j \partial x_k} = \frac{\partial}{\partial x_i}\left(-\frac{\partial \epsilon_{jk}}{\partial x_i} + \frac{\partial \epsilon_{ik}}{\partial x_j} + \frac{\partial \epsilon_{ij}}{\partial x_k}\right) \quad i \neq j \neq k \tag{3.13b}$$

The constitutive law governs the relationship between the stresses and the strains. Although stress and strain are second order tensors, it is convenient to flatten them into vectors so that the fourth order stiffness tensor can be expressed as a matrix. After flattening

these tensors and removing the elements that are redundant due to symmetry the resulting vectors are $\boldsymbol{\sigma} = [\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{23}, \sigma_{13}, \sigma_{12}]^{\mathrm{T}}$ and $\boldsymbol{\epsilon} = [\epsilon_{11}, \epsilon_{22}, \epsilon_{33}, \epsilon_{23}, \epsilon_{13}, \epsilon_{12}]^{\mathrm{T}}$. Equation (3.14) shows the resulting linear constitutive law in which C is the stiffness matrix.

$$\boldsymbol{\sigma} = C\boldsymbol{\epsilon} \tag{3.14}$$

In general, the only constraint on this 6×6 matrix is its symmetry. This results in up to 21 independent elements. However, there are important classes of materials with more constraints and simpler stiffness matrices. Isotropic materials like some metals have the same stiffness in every direction. Their stiffness matrix is defined with just two parameters: Young's modulus E and the Poisson ratio ν. Equation (3.15) shows the isotropic stiffness matrix.

$$C = \frac{E}{(1+\nu)(1-2\mu)} \begin{bmatrix} 1-\nu & \nu & \nu & 0 & 0 & 0 \\ & 1-\nu & \nu & 0 & 0 & 0 \\ & & 1-\nu & 0 & 0 & 0 \\ & & & 1-2\nu & 0 & 0 \\ & \text{sym} & & & 1-2\nu & 0 \\ & & & & & 1-2\nu \end{bmatrix} \tag{3.15}$$

Orthotropic materials like composite plies and balanced composite laminates have two orthogonal planes of symmetry and have a stiffness matrix defined by nine parameters. The

orthotropic stiffness matrix is shown in equation (3.16).

$$C = \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ & C_{22} & C_{23} & 0 & 0 & 0 \\ & & C_{33} & 0 & 0 & 0 \\ & & & C_{44} & 0 & 0 \\ & \text{sym} & & & C_{55} & 0 \\ & & & & & C_{66} \end{bmatrix} \tag{3.16}$$

The governing equations have no closed form solutions for arbitrary structural shapes so approximate solution methods are used instead. The finite element method [23] is the standard approach. The finite element method discretizes the structure into a mesh of elements with a predetermined set of basis functions that relate displacement at the element's nodes to strains within the element. This allows for an approximate solution to the global deformation to be composed from simple deformations within each element. In the linear static case, the finite element method results in a linear equation that maps nodal displacements to nodal forces through a sparse global stiffness matrix. The displacements can then be calculated by a sparse linear solver.

The finite element method can be thought of as a general framework for solving problems, and it allows a great deal of flexibility in how individual structures are modeled. Models do not necessarily need to be a faithful recreation of the actual structure to predict its behavior. Many aerospace structures are panel-like, with a small thickness in one direction, and can be accurately represented by surfaces of two-dimensional finite elements. Composite laminates, composed of individual composite plies that have different stiffness properties because they are oriented in different directions, can be homogenized or "smeared" into an equivalent laminate stiffness [63, 64]. Naked panels are prone to buckling under compressive loads, so realistic structures have stiffeners to add flexural rigidity and these can also be homogenized into an equivalent overall panel stiffness [31]. Smeared panel representations

are advantageous because they significantly reduce computational costs for only a small reduction in accuracy, and they allow the same geometry and mesh to represent structural alternatives with different composite laminates and stiffener dimensions by adjusting the model's stiffnesses.

Aeroelasticity

In order to account for the interaction between aerodynamics and solid mechanics, the aerodynamic vortex lattice model and the structural finite element model must be connected to each other. These two models already give rise to matrix equations; their interaction can be described by additional matrices that capture the effects of the structure's displacement on the normal flow, and the effects of the vortex strengths on the nodal forces. The single-discipline equations and the interaction equations can then be assembled into an overall matrix equation governing the multidisciplinary system, which is then solved to determine the static aeroelastic response.

An important step in developing the interaction matrices is accounting for the different meshes used in the two models. This can happen, for instance, if the structural model only consists of the wingbox (between the front and rear spar), while the aerodynamic model must still extend from the leading edge to the trailing edge. There is also a tendency for the vortex lattice model to have a coarser mesh. The different meshes can be handled through an interpolation scheme that both determines displacements at the vortex lattice control points and also maps aerodynamic forces to the structural model. An example of such an interpolation scheme is the infinite plate spline [65].

Structural Failure

A structure should not fail due to plausible foreseeable operating conditions. Failure theories have been developed to mathematically express the circumstances that cause structures to break. These theories often relate the internal forces or deformation to the material's strength

or stiffness properties to determine whether a set of loads will cause failure. Some of the most important types of failure to consider for aerospace structures are material strength and structural stability. Strength failure theories, like the Tsai-Wu [66] theory for composite plies, predict the stresses or strains that cause the structural material itself to break. Stability failures [67, 68, 69], like panel buckling, local buckling between stiffeners, and crippling, occur when the component suddenly deforms and is no longer able to carry additional load. This could lead to its own collapse or concentrate load in nearby components.

Early design weight estimation models do not account for every potential failure mechanism and every load case. This is because some types of failure, like flutter, are disproportionately computationally expensive to check, and because there are too many load cases to consider at this stage of design. Instead, these weight estimation models try to identify the critical load cases and failures that are most constraining on the structural design. Even if an excluded failure/load case actually would be critical for a portion of the structure, if the included ones impose similar constraints and the weight difference is small, then the exclusion is justified.

Derivatives

It is often useful to be able to obtain derivatives from the physics-based analysis. Ordinarily, this would be for the purpose of applying a gradient-based optimization algorithm [70] to search for the best design. In this case, derivatives are required by the proposed active subspace method.

One of the simplest ways to obtain derivatives is through finite differences [71, 72], in which the analysis is repeatedly run with slightly perturbed input variables. The advantage of this technique is that the analysis program can be used without any modifications, but the disadvantage is that it must be run once for each input variable which can be costly when there are many input variables. Additionally, for the kinds of analysis examined in this chapter, the costs are dominated by factoring the matrix describing the linear system

of equations. With finite differencing, this factorization would need to be recalculated for each perturbation, but with an analytical approach to differentiation it would only need to be calculated once.

Algorithmic differentiation [73] numerically applies the chain rule to each calculation performed by the original analysis. Because of this, it is more accurate than finite differences and can take advantage of cost-saving opportunities like reusing matrix factors. Additionally, algorithmic differentiation can apply the chain rule forwards or in reverse. In forward mode the cost is proportional to the number of input variables and in reverse mode the cost is proportional to the number of output variables. Reverse mode is highly advantageous when there are many input variables used to calculate a single output. For physics-based structural weight estimation, there are a lot of output variables representing the failure constraints discretized over the entire structure, but it is possible to aggregate these constraints [74] to reduce the cost of reverse mode. Despite these advantages, algorithmic differentiation can be difficult to use because it requires access to the analysis program's source code.

3.3.2 Implementation

The testbed follows a standard approach to physics-based structural weight estimation. Given a case's parameters, it starts by creating the planform geometry and structural geometry. These geometries are then meshed so they can be used in an aerodynamic vortex lattice model and a structural finite element model, respectively. Properties are applied to the structural elements and a surface spline is introduced to link the models together. Next, the external loads and the boundary conditions are applied. Finally, design variables and responses are defined for the purposes of calculating derivatives before performing the analysis and obtaining the system responses.

At the top level, the testbed is programmed in Python [75]. After the data defining the model has been created the actual analysis is performed with MSC Nastran [76], a widely used commercial finite-element-based analysis program. In addition to structural analysis,

Nastran has some capabilities for multidisciplinary analysis and optimization, including vortex lattice aerodynamic models. This is beneficial because the entire model can be defined in a single file and analyzed in a single execution. Nastran is called as an executable, so the testbed uses pyNastran [77] to help prepare the text input file and read binary output files.

Geometry and Mesh

The geometry for the testbed is a right half-span wing and tail similar in size to the Common Research Model (CRM) [78], a configuration representative of wide-body aircraft like the 777 and originally intended to serve as a common test case for computational aerodynamic methods. For simplicity, the testbed has no taper or sweep, and as a result the wing planform is rectangular and the wingbox is a rectangular prism. Because the dependence of technology impact on the planform is a focus of the thesis methodology, the planform geometry is parameterized.

The structural layout is a basic two-spar layout with parameters based on the Undeformed CRM (uCRM) [79], a version of the CRM that was adapted for use in structural problems by adding an internal structure. The structural geometry only includes the wing's primary structure or wingbox, consisting of the two spars and the parts of the skin and ribs between the spars. A rib is placed directly at the side-of-body butt line and the rest of the ribs are evenly spaced to match their nominal spacing as closely as possible. A simple flat surface is used for the horizontal tail structure because although a tail structure is needed to receive trim loads from the aerodynamic model's tail, it is not otherwise of interest and does not need to be realistic. Key weight and geometry parameters of the testbed's baseline configuration are displayed in table 3.2.

The geometry is then meshed, and since the planform and wingbox have such simple shapes it ends up regularly ordered and composed entirely of quadrangles. The aerodynamic mesh spacing is chosen so that there are ten elements from the wing's leading edge to its

Table 3.2: Baseline weight and geometry of simple physics-based testbed

Category	Parameter	Description	Value	
Weight	$TOGW$	takeoff gross weight	655,880	lb
	$MZFW$	maximum zero-fuel weight	429,990	lb
	W_{wing}	wing weight	66,770	lb
Wing	S_{wing}	planform area	594,720	in^2
	AR_{wing}	aspect ratio	9	
	b_{wing}	span	2313.5	in
	λ_{wing}	taper ratio	1	
	${}^{t}/_{c\ \text{wing}}$	thickness-to-chord ratio	0.1542	
	$\Lambda_{c/4\ \text{wing}}$	quarter-chord sweep	0	deg
	$\Delta\alpha_{\text{geom}}$	washout angle	5	deg
Horizontal Tail	S_{tail}	planform area	144,000	in^2
	AR_{tail}	aspect ratio	4.9	
	b_{tail}	span	840	in
	λ_{tail}	taper ratio	1	
	$\Lambda_{c/4\ \text{tail}}$	quarter-chord sweep	0	deg
Structural layout	${}^{x}/_{c\ \text{front spar}}$	front spar chord fraction	.1	
	${}^{x}/_{c\ \text{rear spar}}$	rear spar chord fraction	.6	
	Δy_{rib}	nominal rib spacing	28.8	in
	$b_{\text{side body}}$	side-of-body span	115.675	in

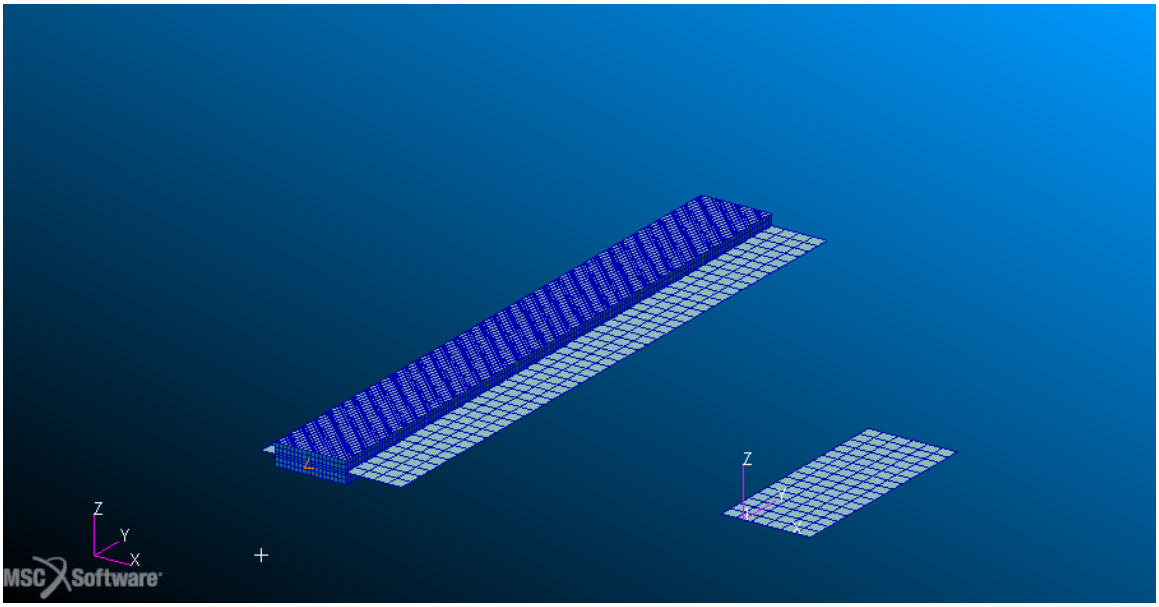

Figure 3.1: Baseline geometry of simple physics-based testbed. Both the rectangular prism finite element model of the structural wingbox and the flat vortex lattice models for the wing and horizontal tail are shown.

trailing edge. The structural mesh spacing is chosen so that there are four spanwise elements between each rib, resulting in around 10,000 elements for the baseline configuration. A visualization of the vortex lattice and finite element meshes is shown in figure 3.1.

The geometry is created using Open CASCADE Technology [80] CAD software, and the structural mesh is generated with SMESH [81]. Since these libraries are not written in Python, the pyOCCT [82] binding was used.

Materials and Structural Properties

The composite ply properties are the same as those used in Brooks, Martins, and Kennedy [41], another study on tow-steered composite structures. The skins are made of unidirectional E752LT/AS4 tape arranged in laminates with ply percentages of 62.5% 0°, 25% ±45°, and 12.5% 90°. The properties of this material are listed in table 3.3. The skins are tow-steered, so these laminate ply angles are all relative to the local steering angle. Note that the majority of plies in the laminate are 0° plies, so will be aligned directly with the steering angle. The

Table 3.3: Mechanical properties for unidirectional tape E752LT/AS4 [41]

Parameter	Description	Value
E_1	longitudinal modulus	17.1 Msi
E_2	transverse modulus	1.41 Msi
ν	Poisson ratio	.34
G_{12}	longitudinal-transverse shear modulus	0.696 Msi
G_{13}	lateral-longitudinal shear modulus	0.696 Msi
G_{23}	lateral-transverse shear modulus	0.696 Msi
ρ	density	0.0560 lb/in^3
X_{t}	longitudinal tensile strength	239 ksi
X_{c}	longitudinal compressive strength	150 ksi
Y_{t}	transverse tensile strength	9.28 ksi
Y_{c}	transverse compressive strength	33.1 ksi
S	shear strength	10.3 ksi

steering angle is defined to rotate about an upwards-pointing axis and so that at an angle of 0° the tows point spanwise.

The spars and ribs are made from a plain weave AS4/8552 fabric, with properties listed in table 3.4. The laminate used is 25% 0°/90° and 75% ±45°. These components are not tow-steered.

For simplicity, no stiffeners are used. Although it would be more realistic, since stiffeners provide more efficient resistance to stability failures, it would also require homogenizing the panel stiffnesses when creating the model and distributing the internal shell forces to the different panel objects to evaluate failure theories. In line with this simplification, only a strength-based maximum strain criteria is used to check for failures in each ply.

Nastran does, however, provides built-in support for composite laminate homogenization. PCOMP bulk data cards allow the laminate stiffness to be defined in terms of the individual plies, with the resulting loads automatically transferred back to the plies so that their failure

Table 3.4: Mechanical properties for plain weave fabric AS4/8552 [41]

Parameter	Description	Value
E_1	longitudinal modulus	9.01 Msi
E_2	transverse modulus	9.01 Msi
ν	Poisson ratio	.045
G_{12}	longitudinal-transverse shear modulus	0.725 Msi
G_{13}	lateral-longitudinal shear modulus	0.696 Msi
G_{23}	lateral-transverse shear modulus	0.696 Msi
ρ	density	0.0560 lb/in^3
X_{t}	longitudinal tensile strength	40.5 ksi
X_{c}	longitudinal compressive strength	38.6 ksi
Y_{t}	transverse tensile strength	40.5 ksi
Y_{c}	transverse compressive strength	38.6 ksi
S	shear strength	10.2 ksi

margins can be calculated. For efficiency, the core of the laminates were homogenized into a single effective ply. However, since the laminate's maximum strains occur on either the top or bottom surface, the outermost plies in each direction were kept distinct.

Loads and Boundary Conditions

The wing's aerodynamic surface is given 5° of washout in order to move the lift distribution inboard during positive-g maneuvers. This helps to counter the large tip loads caused by the unrealistically large tip chord associated with the untapered planform. The aerodynamic loads are transferred from the vortex lattice model to the upper skin via an infinite plate spline, and the spline reciprocally transfers displacements back to the aerodynamic model. The spline is given an attachment flexibility of 1×10^{-4} times the wing planform area to smooth the interpolation.

The mass of the rest of the aircraft is represented by concentrated masses that are

uniformly distributed among the inboard ribs. The fuselage station of the aircraft's center of mass is assumed to coincide with the wing's quarter chord. Interpolation constraint elements (RBE3s) are used to spread the mass over each rib.

The structure is symmetrically constrained at the root rib. Each mesh node on the root rib is prevented from translating along the y-axis, or rotating about the x- or z-axis. The other reactions are applied to a reference node, which is connected to the root rib nodes via an interpolation constraint element so that the forces are evenly distributed. The reference node is constrained in x-translation and fictitiously supported in z-translation and y-rotation, allowing for constant accelerations due to aerodynamic forces. Finally, the tail is rigidly attached to the reference node.

Only one load case is applied—a 2.5g pull-up maneuver at a lift coefficient of 1.25. For the baseline configuration a dynamic pressure of 1.22 psi is used, and in other cases the dynamic pressure is modified based on the configuration's wing reference area to maintain the lift coefficient. Compressibility of the flow is not accounted for. The aerodynamic model is trimmed to the desired load factor by adjusting the pitch of the horizontal stabilizer.

Derivatives

Nastran's solution 200 [83] is used for optimization and design sensitivity. The testbed defines design variables that represent the coefficients of seven spanwise B-splines that control: upper skin thickness, lower skin thickness, front spar thickness, rear spar thickness, rib thickness, upper skin tow orientation, and lower skin tow orientation. Each spline is a third-order cardinal B-spline, meaning that it is a piecewise cubic function and that its knots are evenly spaced. The testbed defines responses for the maximum strain failure margin in each element as well as the total structural weight.

Nastran calculates the derivative of these responses with respect to the design variables using a semi-analytic approach. This approach centers on the solution of the main matrix equation. Finite differences are used to propagate derivatives from the inputs up to this

operation, and from immediately after this operation to the outputs, but analytic differentiation is used for solving the linear system itself. This allows for the matrix factors to be reused without recalculating them. Overall, this semi-analytic approach combines the flexibility of finite differences with the efficiency of analytic differentiation by using analytic differentiation on the particular step where it greatly saves costs.

Nastran can optimize models on its own, in which case the derivatives are only needed for internal use. However, the derivatives must be exported from Nastran to use them in algorithm 1. Although pyNastran was able to read many of Nastran's binary output tables, it was not able to read the derivative table. Therefore, a function was written to read the binary derivative output table according to Nastran's specified format.

Optimization Benchmarks

Although the testbed was created to test the proposed active subspace method, it is also possible to optimize it. Several informal optimization studies were carried out to check the behavior of the testbed and ensure it works as expected. Additionally, optimization will be needed as a baseline to determine how well the active subspace method performs. The objective of the optimization was to minimize the weight of the structure. The maximum strain failure theory was used as a constraint, with a required safety factor of 1.5, and a minimum gauge thickness of 0.02 inches was also required. Each constraint applied to every finite element in the wingbox. The minimum thickness constraints were linear while the structural failure constraints were nonlinear.

Figure 3.2 shows how the optimized structural weight decreases as the number of spline coefficients increases. By the time there are 19 coefficients per spline, corresponding to 16 piecewise-cubic polynomials, there is hardly any improvement possible from adding more coefficients, indicating convergence. It also shows that there is a large weight penalty from using a low-dimensional polynomial parameterization, which supports the need to find an active subspace within the high-dimensional parameterization rather than use a naive

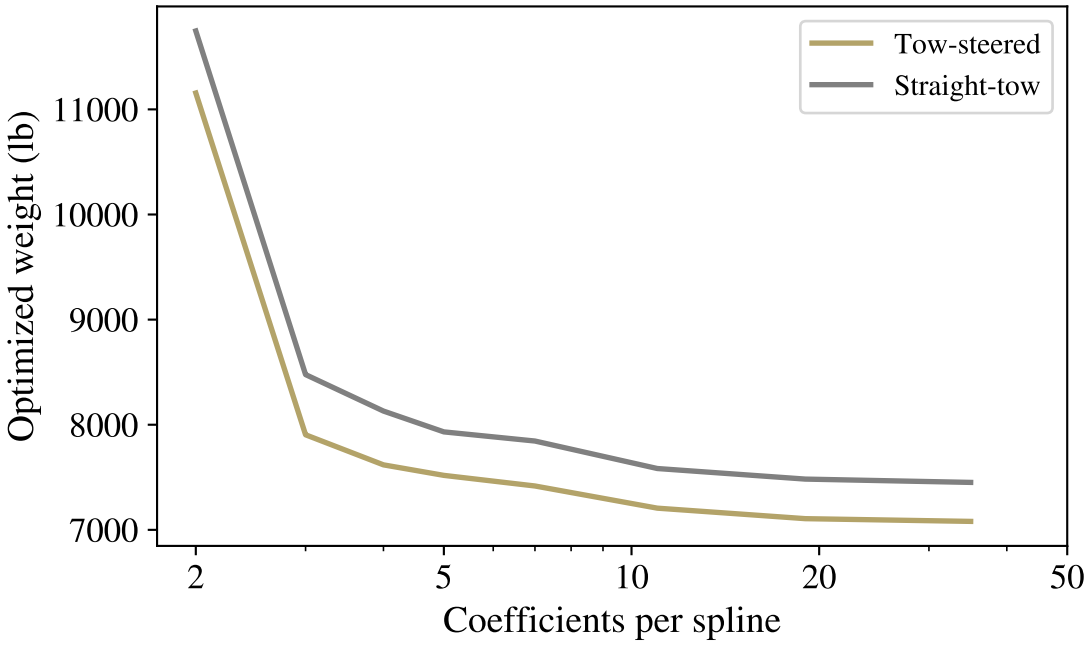

Figure 3.2: Convergence of the optimal weight as the number of B-spline coefficients increases

low-dimensional parameterization. Figures 3.3 and 3.4 visualized the optimized baseline configuration's stresses and failure margins, respectively.

Figure 3.5 shows how the optimal weight converges as only the number of tow path design variables increases, with the number of thickness variables held at a high constant value. Note that the y-axis of this figure spans a much smaller range than that of figure 3.2, indicating that the tow paths have smaller effect than the thicknesses. A large amount of the benefit provided by tow steering comes from just the first coefficient, but there are still noticeable improvements up to around ten.

As mentioned, Nastran has optimization algorithms and the capability to optimize models internally. However, during the course of these tests these algorithms were found to have difficulty with the tow-steered composite wingbox. The most promising ones were a modified method of feasible directions algorithm, which tended to unnecessarily stick to constraints, and a sequential unconstrained minimization technique algorithm that was

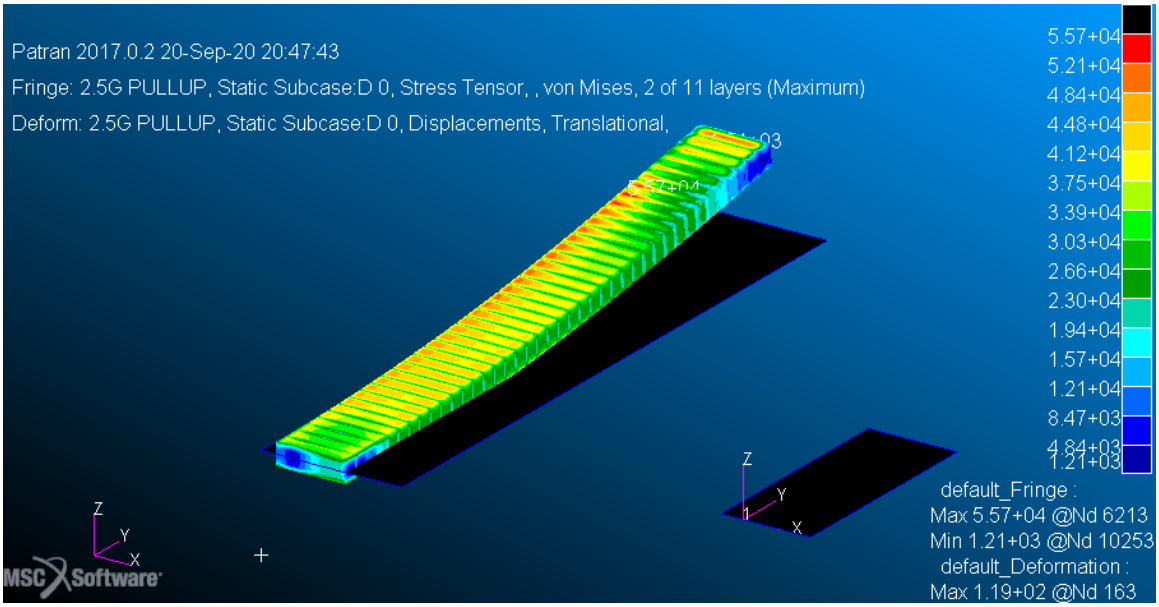

Figure 3.3: Stress of optimized baseline simple physics-based testbed. The displayed stresses are von Mises stresses of the averaged overall laminate. The deformation of the structure is also shown in true scale.

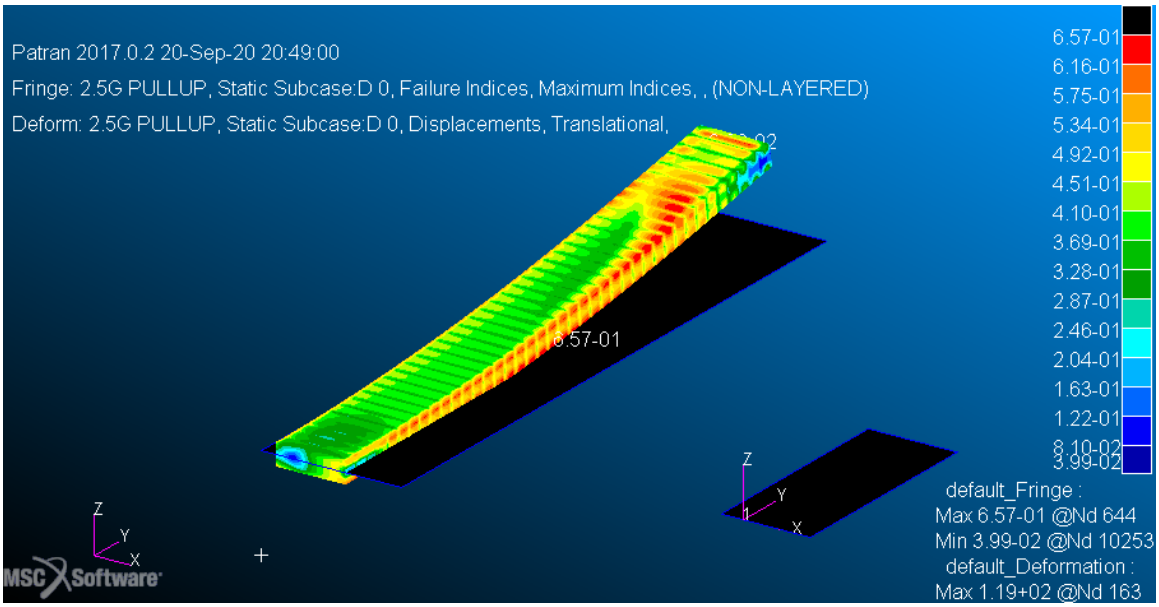

Figure 3.4: Failure margins of optimized baseline simple physics-based testbed. The failure margins are based on a maximum ply strain failure theory. The deformation of the structure is also shown in true scale.

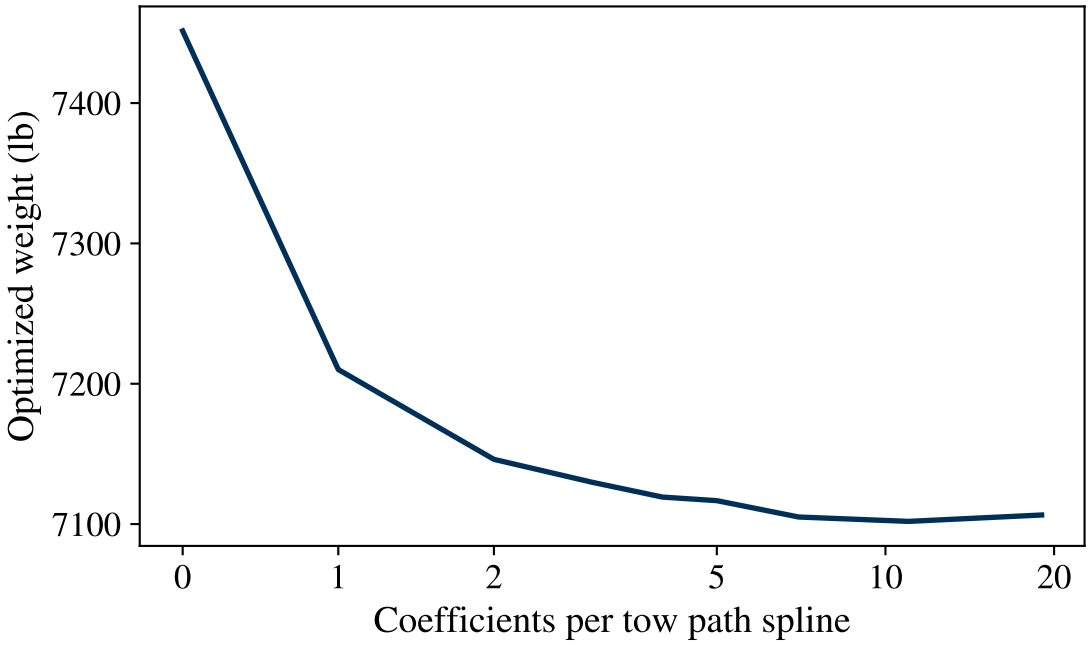

Figure 3.5: Convergence of the optimal weight as the number of tow path coefficients increases and a constant large number of thickness coefficients

inaccurate and slow to converge. Using an external optimization algorithm added significant amounts of input/output time due to repeatedly calling the Nastran executable and exporting the derivatives. Despite this, it was found that SciPy's [84] implementation of SLSQP [85], a sequential quadratic programming algorithm, was able to find a better quality solution much more consistently than Nastran's built-in algorithms. This superior performance is almost certainly due in part to SLSQP updating a curvature estimate with every iteration. In contrast, Nastran's internal optimization approach learns no curvature information between major iterations regardless of the selected algorithm.

Effects of Simplifications

Most real wingbox structures are semi-monocoque, meaning that the skin is reinforced by stiffeners but still bears a significant portion of the load itself. The fact that this testbed uses unstiffened panels means that it represent a monocoque structure in which the skin supports

all of the load. This type of structure was chosen to simplify development of the testbed, since Nastran does not provide built-in support for homogenized stiffened panel structures or analytical buckling failure analysis. Either third-party software or additional code would be needed to homogenize the panel stiffnesses, split the internal forces calculated by finite element analysis into separate panel and stiffener loads, and implement analytical buckling analyses.

This simplification has several important consequences. First, since stability-based structural failures are not considered, the sized testbed structure should have a lower weight than a structure sized to resist stability-based failures. Second, since the testbed does not have stiffeners and does not need to parameterize their dimensions, the testbed has fewer design variables. There are also some consequences that are specific to tow steering. Brooks, Martins, and Kennedy [41] found that one of the major sources of tow steering's weight reduction comes from improved resistance to local panel buckling. Since this testbed does not account for local panel buckling, the benefit from tow steering should decrease. On the other hand, since all of the load is carried by the tow-steered skin rather than by stiffeners, the testbed will benefit more from improved resistance to material strength failures, and the benefit from tow steering should increase. If the effect on local panel buckling is larger then the testbed should underestimate the technology benefit, but this is not completely clear.

Another important difference between this testbed and a realistic CRM-like aircraft's wing is the lack of sweep. Sweep is necessary to prevent shock waves and the resulting drag increase from appearing in the transonic flight regime. Holding wingspan constant, sweep increase the length of the wing's equivalent beam axis, and so increases the root bending moment. Sweep also moves the wing's center of pressure back, which causes a larger root twisting moment. Tow steering might be able to adjust the load paths to improve the structure's ability to cope with these moments, so the lack of sweep in the testbed could reduce the potential benefit from tow steering.

The limitations of this testbed are somewhat similar to those of the analytical structural

weight estimation methods described in section 2.2, which also only account for strength-based failures and the amount structural of material. Despite their limitations, these methods have proven to be useful tools for approximating structural weight. Their success as first-order approximations indicates that even with the missing details, these methods are able explain a large majority of the structure's weight. It therefore stands to reason that if the proposed active subspace method is successful on the testbed, it would also be likely to succeed if stiffeners and stability failures were modeled.

3.4 Experiment 1

There are several questions about the proposed active subspace method given by algorithm 1 that need to be addressed via experimentation. Most importantly, can this method be used to find the active subspace of the tow-steered composite wingbox design problem? And if so, is that active subspace small enough and does it explain enough of the behavior to permit surrogate modeling without sacrificing performance? If not, then the method will not be an effective approach to the overall problem and hypothesis 1 should be rejected. Additionally, there are several parameters of the method that could affect its success, but the size of their effects and good settings for them are not known. Learning that information would also shed light on the whether the benefits of the augmented Lagrangian formulation are worth its added complexity relative to the quadratic penalty.

3.4.1 Procedure

An experiment was needed in order to find the effects of the method's parameters and ultimately to determine if the method is successful. The experimental factors were the method parameters: the curvature, the penalty parameter, and the trust region radius, plus an additional factor for the number of retained constraints.

The number of retained constraints was studied because the optimization problem has a very large number of constraints due to there being a failure constraint associated with

Table 3.5: Factor settings for active subspace experiment

Factor	Settings		
	Low	**Medium**	**High**
curvature	0.2	0.632	2
penalty parameter	0.1	0.316	1
trust region radius	$0.2\sqrt{n}$	$0.632\sqrt{n}$	$2\sqrt{n}$
retained constraints	1	3	10

each finite element. This increases the difficulty and expense of the Lagrange multiplier estimation subproblem. Since the failure margins in nearby elements should be correlated, it makes sense to screen the constraints [86] to reduce the redundancy. Constraint screening is implemented by grouping the constraints by face (e.g., all the elements on a single rib or all the elements on a single skin panel) and keeping only the worst constraints from each group.

The experimental factors and their ranges are shown in table 3.5. A face-centered central composite design of experiments was selected in order to capture the overall trends and allow for the possibility of curvature and two-factor interactions. This design has a total of 25 cases. Since each factor scales an aspect of the method, they were all put on a logarithmic scale. Each factor was given a range of one decade.

The configuration under consideration was the simple physics-based testbed described in section 3.3. Framing the overall goal as a parameterized optimization problem, the parameters were the planform's wing area and aspect ratio, and the choice/design variables were the thicknesses and tow orientations. Each of the thickness and orientation B-splines was third order and had nineteen coefficients. This corresponds to sixteen smoothly joined cubic polynomials, and resulted in a total of 133 design variables.

The only response tracked was the optimal weight of the baseline planform configuration. This was obtained by finding the active subspace with each run's factor settings, and then searching for the optimum in that subspace with a gradient-based algorithm. The thought is

that, if a good active subspace has been found, then the optimum in the lower-dimensional space should closely match that of the original high-dimensional problem. The quality of the active subspace is inferred from the optimized weight associated with it. The mean of the weighting distribution was used as the origin of the reduced-dimensional space.

Rather than manually examine the eigenvalues for the presence of a gap within each run, it was assumed that the active subspace would have 21 dimensions. A high-level goal of this effort is to enable surrogate modeling, and 21 dimensions is near the upper limit of the amount that a surrogate model approach can handle. The number 21 was chosen in particular because there are 7 design variable splines, so choosing a multiple thereof allows for a direct comparison between the low-dimensional active subspace parameterization and a low-dimensional polynomial parameterization.

Weighting Function

A weighting distribution over the input space was needed to perform the method. This distribution corresponds to a region of interest for the function and controls where the gradient samples are draw from. The size of the distribution reflects the size of the design space. A wide distribution should lead to an active subspace that is applicable over that wider space, while a tight distribution should be more likely to lead to a smaller active subspace since it sees a smaller portion of the function (in the limit with a very concentrated distribution the function is approximately linear and the active subspace has only one dimension).

The weight distribution was obtained by solving a preliminary set of optimization problems over a range of planform settings and fitting probability distributions to the population of optimal design variables. This is in a sense cheating by using a known answer to help find the solution via a different method. To mitigate this, all design variables were pooled together into a thickness or a tow orientation group. The hope was that this pooling would create a representative distribution for the design variables without narrowing the

Table 3.6: Input variable marginal distributions

Variable	Distribution	
wing area	$\mathrm{Uniform}(a=450{,}000, b=750{,}000)$	in^2
aspect ratio	$\mathrm{Uniform}(a=9, b=13.5)$	
thickness	$\mathrm{Lognormal}(\mu=-2.19, \sigma^2=2.04)$	in
tow orientation	$\mathrm{Normal}(\mu=7.17, \sigma^2=102)$	deg

range by an unrealistically prescient amount. As a result of this, the middle of the distribution is below the optimal inboard thicknesses and above the optimal outboard thicknesses. The marginal weight distribution for each input variable is shown in table 3.6.

This weight distribution was also used to scale the input variables. From the active subspace procedure's point of view, each input distribution had a standard deviation of one.

Sampling

The recommended number of samples for finding the active subspace is $N = \alpha k \log n$, with number of original dimensions n, target reduced number of dimensions k, and oversampling factor α. The experiment has 133 design variable dimensions that will ideally be reduced to only 21 active subspace directions. An oversampling factor of 10 was used, resulting in a total recommended number of samples of 1,027. This many random samples were drawn from the input distribution, and for each set of inputs the testbed was executed to determine the value and design sensitivity of both the objective and the constraint functions.

Executing the testbed 1,027 times is a large computational expense, so for efficiency only one set of samples was generated and was reused in each of the 25 runs of the experiment. However, since the experimental factors affect the Lagrange multiplier estimates, the multipliers needed to be recomputed based on the factor settings in each case.

Gradient Calculation

Before estimating the Lagrange multipliers and calculating the gradient of the augmented Lagrangian, several of the problem's quantities were scaled and preprocessed. The objective function, structural weight, is a linear function of the thickness coefficients. However, the gradient of this function is larger in the optimization problems corresponding to larger planform areas. In order to avoid the cases with larger planform area having a disproportionate affect on the active subspace the objective function was scaled so that the typical size of the elements of its gradient would be one. The constraints were also normalized in a similar fashion.

A particular form of maximum strain failure constraint was used in the procedure. This kind of constraint is commonly expressed using the form shown in equation (3.17a), in which the strain ϵ is divided by the ultimate strain $\epsilon_{\text{ultimate}}$. This usually keeps the value of the constraint at $\sim$1 but for small structural dimensions the strain value approaches infinity. Within the procedure this would be difficult to avoid, since the samples are generated randomly, and when it occurred would cause extreme gradients that distort the results. To prevent this, the constraint form in equation (3.17b) is used instead. It should be apparent that both forms reflect the same feasible region, as both equal zero when the strain equals the ultimate strain. The difference is that as the strain approaches infinity the value of the alternate constraint only approaches one. This form is a monotonic transformation of the more common form so it does not affect the constraint screening.

$$g = \frac{\epsilon}{\epsilon_{\text{ultimate}}} - 1 \tag{3.17a}$$

$$g = \frac{\epsilon - \epsilon_{\text{ultimate}}}{\epsilon + \epsilon_{\text{ultimate}}} \tag{3.17b}$$

3.4.2 Results and Discussion

Analysis of Trends

A quadratic response surface model was fit to the data using linear least squares regression in an attempt to learn the trends associated with the experimental factors. The model did not perform well: it had an R^2 of only .61 and had a PRESS (cross-validation) R^2 of -2.23, indicating comically severe overfitting. Another model was fit using adaptive lasso regularization with an AICc validation criteria to avoid overfitting, but this model did not find any of the factors to be useful predictors.

Examining the data revealed an outlier that was potentially the source of this difficulty. The outlier is clearly visible in figure 3.6, which shows a histogram of the optimized structural weight from each case. It is possible that something pathological or fundamentally different happened in that case, resulting in behavior that is too complicated to be captured by a simple second-order effects model.

The outlier was excluded before fitting another lasso-regularized model to the data. This time the model was able to detect significant trends due to the factors and had an R^2 of .915. The significant effects and their sizes are listed in table 3.7. The largest effect was from the trust region radius, with better performance associated with a larger trust region. However there was also a strong curvature effect from the trust region radius, so its most favorable setting is within the experimental range rather than at the extreme. Better performance was also predicted for using a small penalty parameter, for retaining few constraints, and for using a small curvature. There are significant interaction terms involving the trust region radius but these only diminish the main effects; none of them are strong enough to reverse the trends.

Interestingly, the best setting predicted by the model uses a trust region radius of 1.14, which means that in this case the heuristic of scaling the trust region radius by the weighting distribution's standard deviation works well. Also of note is that since the augmented

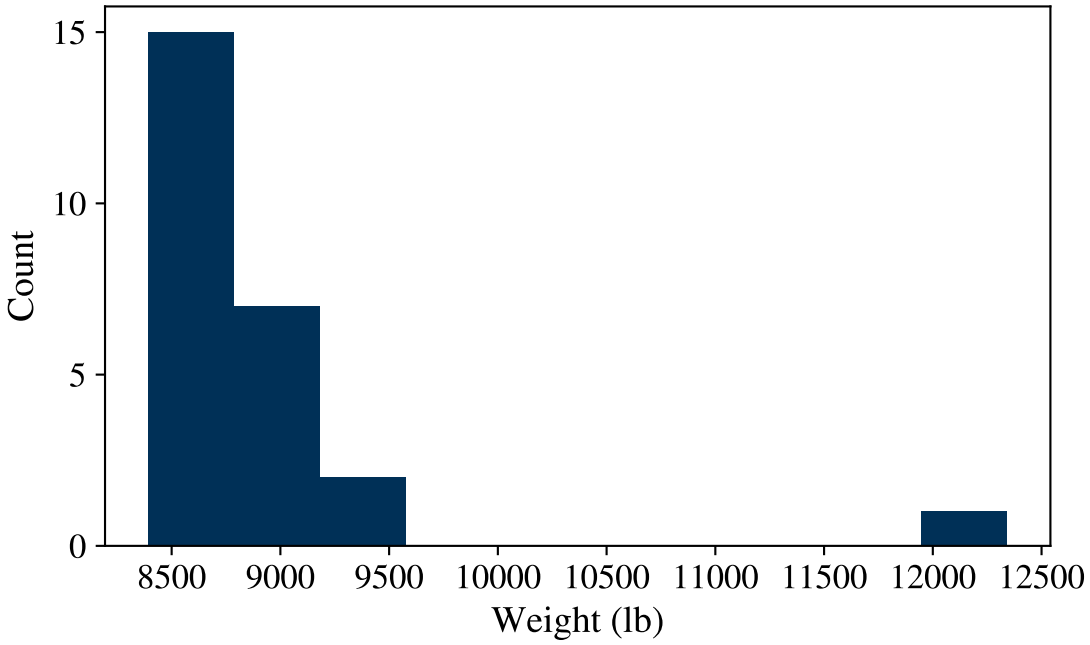

Figure 3.6: Histogram of optimal weights from initial active subspace design of experiments. There is a clear outlier.

Lagrangian aggregation simplifies to quadratic penalty aggregation when the trust region is small and the curvature is low, and because the model suggests that this is not the best setting, there is evidence showing an advantage for the more complicated augmented Lagrangian style aggregation.

Unfortunately, the best settings predicted by the model (high trust region radius, low penalty parameter, low retained constraints, and low curvature) also coincide with the excluded outlier. A follow-up case was run with the penalty parameter, retained constraints, and curvature at their lowest value and with the trust region radius at a high interior value. The response for this case was an optimized weight of 10,400 lbs, which was significantly worse than the model predicted but not as bad as the outlier.

This suggests that the outlier was not a fluke but rather caused by a nonlinear change in behavior that takes over if the trends in table 3.7 are pursued too aggressively. This nonlinearity also means that a second-order effects model is unable to make accurate

Table 3.7: Effects of active subspace method's parameters

Factor	Effect size
trust region radius	-204
penalty parameter	120
retained constraints	88
curvature	84
trust region radius $\times$ trust region radius	212
curvature $\times$ trust region radius	71
penalty parameter $\times$ trust region radius	-58

predictions in this corner of the experimental factor space.

An additional eight follow-up runs with manually selected factor settings were executed to try to find the best possible factor settings, bringing the total number of cases up to 34. All but one of these cases yielded results consistent with the model. The other one was located in the troublesome corner of the experimental space and like the others in that corner performed much worse than predicted by the model. These follow-up cases provided an additional measure of confidence in the model's predictions and interpretation of the trends, at least over the majority of the experimental domain. They also helped to confirm that the model's occasional poor predictive ability was confined to a specific region and that the reason for this is a different behavior and not coincidence.

Analysis of the Best Case

The details of the best case's settings and results are shown in table 3.8. This case had low curvature, a high trust region radius, and a middling penalty parameter and number of retained constraints.

The uncentered covariance matrix of this case's aggregated gradient is shown in figure 3.7. The gradient samples were scaled based on the design variable ranges to prevent any

Table 3.8: Factor settings and results for best-performing active subspace

Variable	Value
curvature	0.2
penalty parameter	0.350
trust region radius	$1.13\sqrt{n}$
retained constraints	5
optimized weight	$8{,}300$ lb

influence from units. The elements of the matrix are ordered by variable group, and from wing root to tip within each variable group. Each group of variable interactions forms a distinctive box in the overall matrix.

The most striking pattern from this matrix is the difference in magnitude between the covariance associated with the thickness variables and that associated with the tow angle variables. This is seen from the much sparser activity in the last two row and column groups. This indicates that even though tow steering affects the structure's performance, the size of its influence is much less than that of the thickness.

Another noticeable pattern is in the interactions between thickness variables. Starting with the boxes along the diagonal, which represent the covariance within a group of variables: the lower and upper skins both have strong positive correlations for the thicknesses at the root and those at the tip, but strong negative correlations between the root and tip. The front spar has a qualitatively similar but weaker covariance pattern. The rear spar and rib thicknesses are positively correlated across the whole wingspan. The root thicknesses are positively correlated between the skins and front spar, but negatively correlated between these groups and the rear spar and ribs.

As for the tow orientations, there is a noticeable diagonal indicating that adjacent sections have positive correlation, but there are no strong correlations between the tow paths at distant sections of the wing.

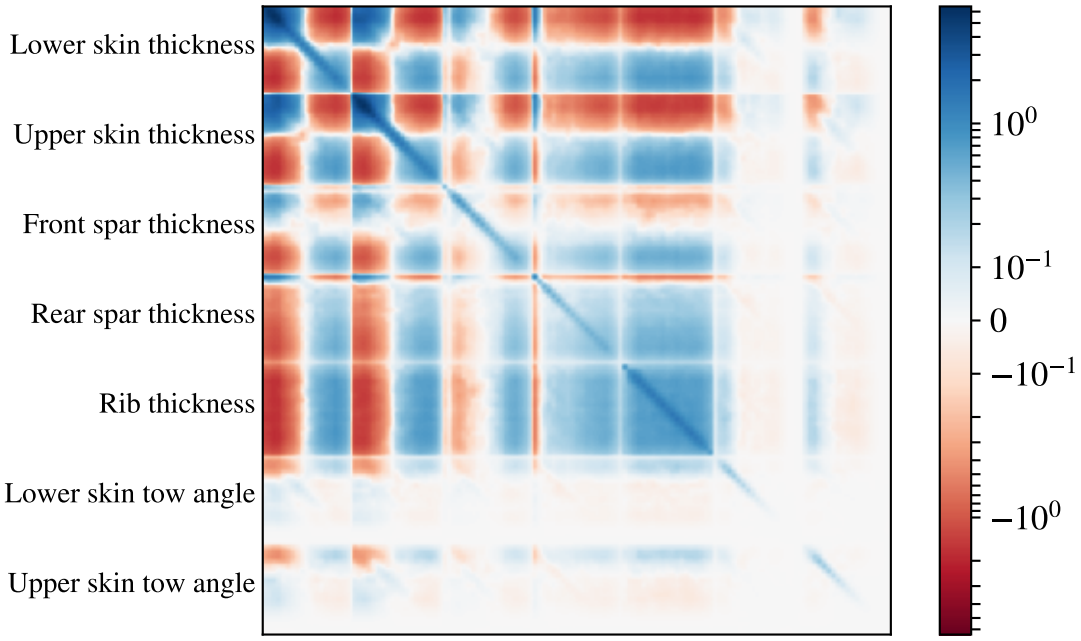

Figure 3.7: Uncentered covariance matrix of the best case. The elements are normalized by the variable ranges and ordered by group, and from tip to span within each group.

The sorted eigenvalues of the uncentered covariance matrix are plotted in figure 3.8. The first couple of eigenvalues are much larger than their immediate follower, but after that there is a prominent shelf with a very slow rate of eigenvalue decay. It is not until the 121^{st} eigenvalue that there is another appreciable drop. This indicates that there is not a low-dimensional active subspace associated with the tow-steered composite design problem, and that only working with the first 21 eigenvectors will probably incur a significant loss of performance compared to the original high-dimensional parameterization. This is also shown through the fraction of variance explained in figure 3.9, which shows that a large number of eigenvectors are needed to fully explain the correlation patterns from figure 3.7. At least 40 eigenvectors are needed to get to 90% of the variance explained, which is likely to be too many for surrogate modeling.

The most important design directions learned by the active subspace procedure correspond to the eigenvectors associated with the largest eigenvectors. Figure 3.10 shows these

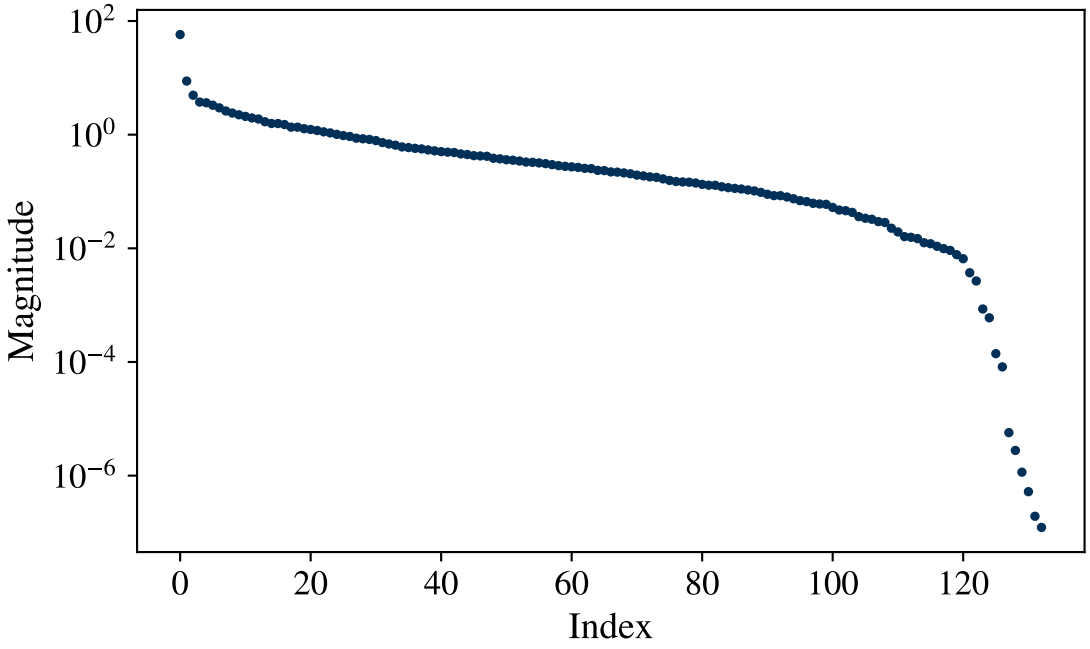

Figure 3.8: Sorted eigenvalues of best case

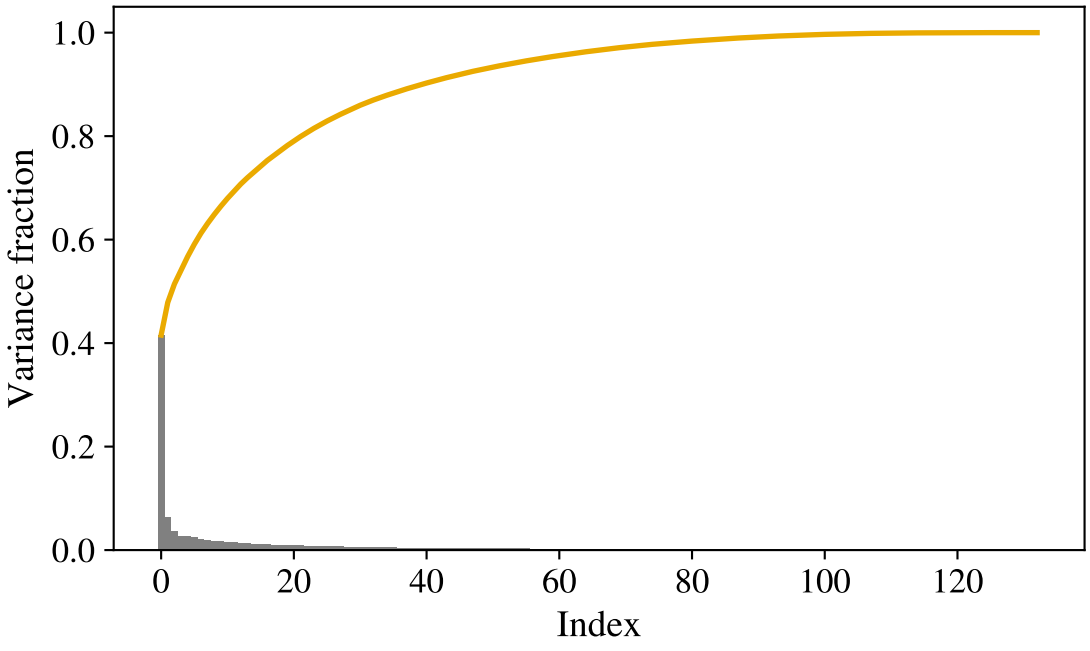

Figure 3.9: Fraction of variance explained by each eigenvector. The gray bars show contributions from individual eigenvectors and the gold line shows the cumulative sum.

directions and demonstrates that this method has been able to learn real features of the problem.

The first eigenvector gives a direction that primarily increases the root skin thickness while decreasing the tip skin and rib thickness. There is a similar but smaller effect on the spars, and a slight adjustment to the inboard tow angles. The shape of this eigenvector fits nicely with physical intuition, since the primary load on the wingbox is a bending moment carried by the skins that is weakest at the tip and strongest at the root. The shape also reflects the need for the spars to bear a shear load that is strongest at the root.

The next couple of eigenvectors are focused on redistributing material between the upper and lower skins. The second eigenvector is situated over the root quarter of the span while the third sits just outboard the second. These two directions also seem intuitively reasonable because the root skins are some of the thickest portions of the structure so it is worthwhile to focus design effort on them. These directions also nicely complement the first eigenvector, which moves the upper and lower skins together, by moving them in opposition.

The eigenvectors after the first three exhibit increasingly high frequency oscillations and are difficult to interpret. These correspond to the shelf of relatively constant eigenvalues in figure 3.8. Since the eigenvalues are close in size it should be expected that the eigenvectors bleed into each other and fail to separate cleanly.

Finally, to assess the success of the procedure developed in this section, the performance of the active subspace from the best case is compared to the original high-dimensional parameterization as well as to a low-dimensional polynomial parameterization. The original high-dimensional parameterization uses seven cubic B-splines and has a total of 133 design variables. The active subspace parameterization uses the first 21 eigenvectors learned using the procedure from this chapter. The low-dimensional polynomial parameterization uses seven quadratic polynomials for a total of 21 parameters as well. The relative performance is shown in figure 3.11. Disappointingly, not only did the active subspace fail to match the performance of the original high-dimensional parameterization, it also performed worse

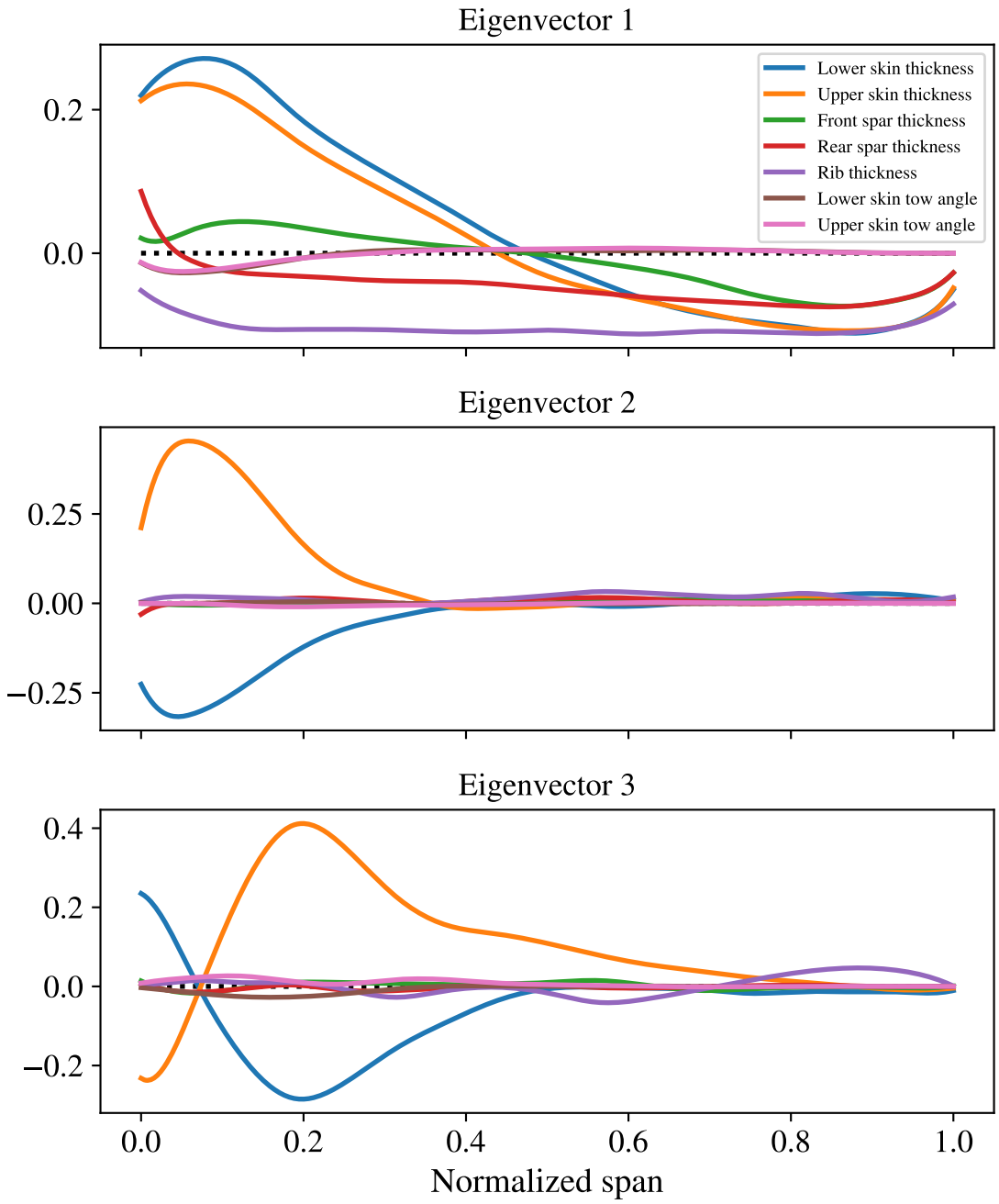

Figure 3.10: First three eigenvectors of best-performing case. The colored lines show the normalized change in the design over the span associated with moving along the eigenvector. The dotted black line shows zero for reference; a value of zero indicates that the design is unchanged by moving along the eigenvector.

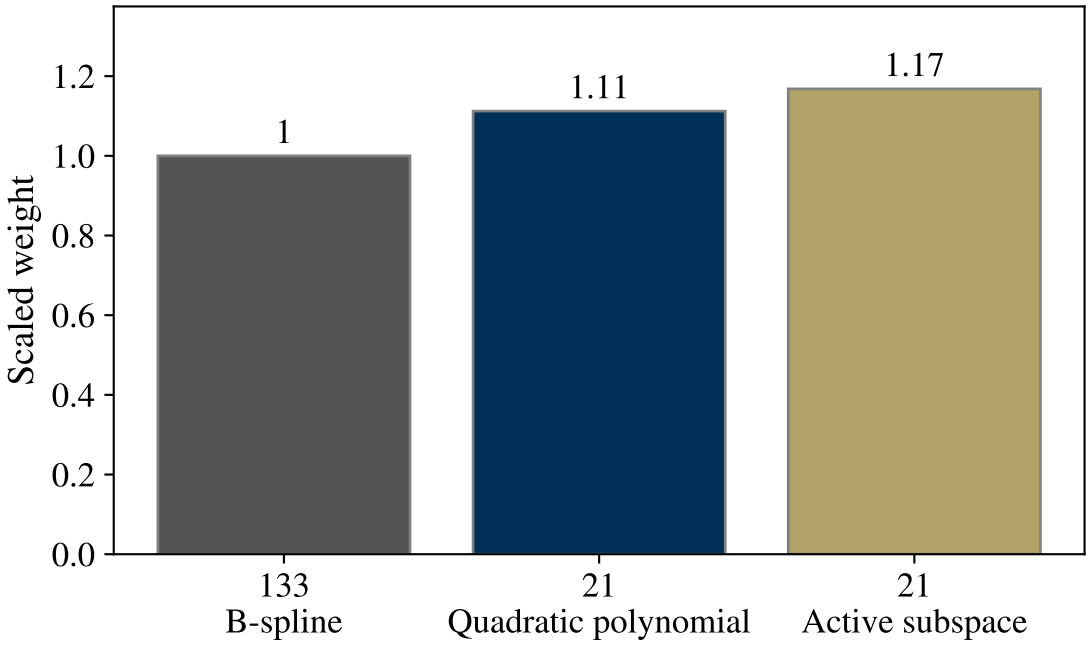

Figure 3.11: Optimized weights of baseline and low-dimensional parameterizations

than the quadratic polynomial parameterization, and was a total of 17% heavier than the original parameterization. This poor performance is likely due to only three eigenvectors being clearly determined: rather than having 21 useful directions to search through the optimization algorithm only had 3.

Conclusion

The method for finding the active subspace of a constrained optimization problem developed in this chapter showed a limited measure of success. It was able to learn a few intuitive design basis directions without being given any specialized knowledge; all it had were samples of the objective and constraint functions and their gradients. Additionally, the method's parameters were shown to have a predictable effect on the method's performance, as was expected.

However, when applied to the tow-steered composite design problem, the procedure did

not perform well enough to meet the stated goals. The active subspace that was identified, consisting of only three directions, did not explain enough of the optimization problem's variation leading to a significant weight penalty. Further, the number of dimensions that would need to be added to practically eliminate this penalty would preclude surrogate modeling. It is not known whether this is due to a shortcoming of the method or if it should be attributed instead to the nature of the tow-steered design problem itself. Finally, the basis directions that were clearly identified as important were almost entirely concerned with the thicknesses. While these directions do appear to reflect fundamental aspects of the structure's behavior, they do not provide knowledge that can help to create a better parameterization for tow path orientations. It is possible that some of the simplifications used by the testbed (no stability failures and no sweep) are part of the reason why the tow path covariance was small and why the tow paths were mostly absent from the clearly identified eigenvectors. A more realistic physics-based model would show increased sensitivity to the tow paths. However, the added realism would also increase the amount of design flexibility needed to obtain a near-optimal design. Since the active subspace method was not able to find an adequate reduced set of design directions for this simpler problem, it does not seem likely that it would succeed on a more realistic problem.

In light of these results, hypothesis 1—that the active subspace could be used as a low-dimensional parameterization suitable for surrogate modeling without sacrificing much performance—must be rejected. However, the thesis methodology still needs a way to formulate the structural design problem and parameterize the tow paths. Since a successful reduced-dimensional approach was not found, as a compromise, the methodology will handle the large number of design variables by decomposing the problem into separate local optimization problems as in fully stressed design. The stiffened panel dimensions will be set in these local problems, while the tow path and planform variables will be included in a surrogate model due to their global effects. This compromise will allow a surrogate model of the physics-based analysis to be created, but will limit the size of the tow path

parameterization and thereby the technology benefit.

CHAPTER 4
SURROGATE MODELING OF PARAMETERIZED OPTIMIZATION PROBLEMS

4.1 Background

There is a difficult trade-off to consider when making a surrogate model of tow-steered composite structural weight. On one hand, as evidenced by figure 3.5, as many input variables as possible should be allocated to defining the tow path, so that the maximum benefit can be obtained. On the other hand, as input variables are added it becomes less likely that the surrogate model will be accurate and at the very least more data will be needed to create it. This second part is important because the physics-based analysis used to obtain each sample is expensive. It is therefore important to carefully plan out the settings used in each analysis run to maximize the efficiency with which useful information is gained, leading to the following gap:

GAP 2
Because the trade-off between technology benefit and computational cost is significant, it must be made as efficient as possible. This creates the need to evaluate the expensive physics-based analysis at carefully selected points that reflect the system's behavior and the experimental goals.

4.1.1 Single-Stage Design of Experiments

Design of experiments refers to a systematic method for selecting which input combinations are examined over the course of an experiment [87]. Many types of designs of experiments have devised, each addressing different experimental goals or systems with different assumed statistical models.

Different experimental goals are reflected in some of the classical criteria used to design experiments. The D-optimality criterion seeks to estimate the coefficients of the statistical

model as accurately as possible, which is useful for inferring the effects each factor has on the response. The I-optimality criterion seeks to minimize the model's prediction variance, which is good for being able to accurately predict the response. The I-optimality criterion is more in line with the goals of the tow-steered composite research, as an accurate surrogate model could be optimized to predict the optimal tow path for a given planform.

Knowledge of the system's behavior is another important consideration for developing an experimental design. Ironically, the design must be chosen when this knowledge is the smallest because that occurs before any cases are run [88]. Classical designs of experiments were developed for physical systems and are best for fitting simple statistical models from a minimal number of samples. The amount of noise in the response is normally an important parameter of the models and the experimental designs are intended to be able to estimate it. Some examples are two-level fractional factorial designs, which are effective for fitting models with linear terms, and central composite designs, which add a center point and axial points to the two-level fractional factorial in order to resolve quadratic terms.

Computer experiments tend to use different designs [89] for several reasons. They are typically easier to run in large quantities than physical experiments, though there is still a need to systematically plan the data collection. Unlike in physical experiments, where there is a possibility that some factors have not been identified or that uncontrolled factors vary systematically, a computer experiment's input factors are perfectly defined so there is no need for blocking. Perhaps most importantly, computer codes can produce nonlinear behavior and, because it is usually possible to make strong assumptions about the amount of noise, the nonlinear behavior is resolvable. Although there can be interactions, it is usually this nonlinearity that is the source of complexity for computer simulations [90]. These general features of computer experiments are applicable to tow-steered composite structure sizing.

For these reasons Gaussian process [91] models have been recommended for computer code responses. Rather than assume a functional relationship between the inputs and output,

they make assumptions about how the response at nearby points in input space are correlated. This gives them the flexibility to represent nonlinear behaviors. More detail on Gaussian processes will be given in section 4.2.1. Space-filling designs of experiments, like maximin Latin hypercubes [92], are recommended to collect the data for these models. These designs spread out sample locations throughout the interior of the experimental space in contrast to classical designs that focus more on the extreme edges. A standard heuristic for the number of samples to collect is ten times the number of input dimensions [93], though ultimately the amount of data required for an accurate model depends on the behavior of the specific system being investigated.

Based on these suggestions, a reasonable strategy for creating a surrogate model for a tow-steered composite structure sizing program is to collect data from a maximin Latin hypercube and model the data as a a Gaussian process. If this results in a sufficiently accurate surrogate then the job is done; the tow paths can be optimized via the surrogate for any planform and the predicted technology benefit can be used in system-level design. However, as will be seen in sections 4.5 and 6, the accuracy of the initial surrogates are less than ideal.

It will therefore be necessary to augment the first stage of data collection. Loeppky, Sacks, and Welch [93] showed that the accuracy converges slowly as larger space-filling designs are used, so the simple strategy of augmenting the initial data with more space-filling points would not be effective. Rather than trying to develop a surrogate model that is accurate over the entire experimental space, it will be necessary to prioritize the regions that best help achieve the experiment's goals.

4.1.2 Parameterized Optimization

In order to properly choose the data to add in follow-up stages, the experimental goals must be precisely defined. The goal for this research is to predict the weight reduction from tow-steered composites as a function of the planform. The system designer wants to

choose the configuration based on overall system-level considerations, and wants an accurate estimate of the technology benefit from using the best tow paths for the chosen configuration. This creates a hierarchy of variables that will be detailed in the rest of this subsection, and as a result this problem is classified as a parameterized optimization problem.

At the top of the hierarchy are the configuration variables such as those defining the wing planform. Although these have a large effect on the structure they must be chosen based on multidisciplinary considerations. For instance, a structure-centered design would minimize weight by selecting a small aspect ratio, but this would lead to high induced drag and poor system performance. This is particularly applicable to tow-steered composites because they have the potential to enable higher aspect ratio wings by reducing the associated weight penalty.

The lower levels of the hierarchy consist of structural design variables. These do not interact much with the other disciplines and as a result can be decoupled and optimized for purely structural considerations. The number of these variable levels depends on the specific design process being used; as a result of the findings in chapter 3 this research will use a fully stressed design type approach and there will be two structural variable levels.

The middle level contains the global structural design variables. These variables have a strong effect throughout the whole wingbox and include the tow paths as well as structural layout variables like spar location, rib spacing, and stiffener spacing. They can be set to minimize structural weight, but this can be challenging for structural layout variables because of the potential for discrete changes in the structural topology or mesh, or for the mesh to become poor from distortion. When a surrogate modeling approach is used, the total number and the ranges of both the configuration and global structural variables need to be managed so that adequate model accuracy is achievable. Depending on the goals of a study some of these variables might be held constant.

At the lowest level are the local structural variables, consisting of panel dimensions like skin thickness, web height, and flange width. These variables primarily affect the portion of

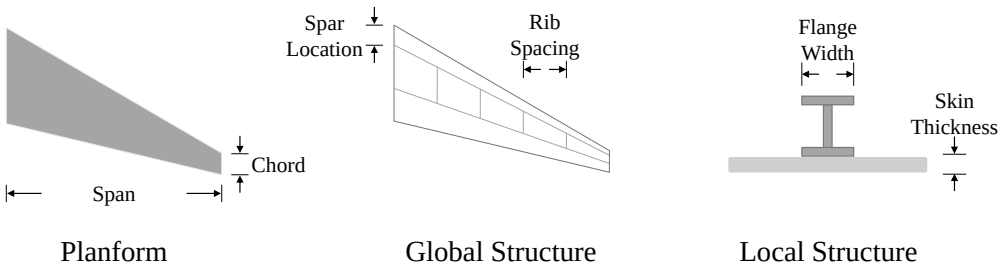

Figure 4.1: Examples of design variables from each level of the hierarchy

the structure they are defined in but also have a smaller effect on neighboring regions. A fully stressed design type approach ignores these interactions but can still produce good, if not optimal, designs. The benefit from this approach is that design can be conducted through many small local weight optimization problems rather a giant global one. For instance, if there are 4 dimensions defining the front and rear spar, upper and lower skin, and ribs across 50 rib bays, then there will be 1,000 design variables. These can be set in 250 4-dimensional optimization problems, which is much easier that a 1,000-dimensional optimization problem. For surrogate modeling, the local structural variable are optimized in each run based on the settings for the configuration and global structural variables. This is a big advantage for surrogate modeling because it allows this large group of variables to be implicitly included without increasing the number of independent variables in the model.

Figure 4.1 summarizes the variable groups and shows examples from each. This variable structure: planform variables that are explored in higher-level design, middle-level global structural variables that are optimized, and low-level local structural variables that are implicitly included through yet a lower level of optimization, forms a family of related optimization problems parameterized by the planform variables. In the economics literature this kind of problem is called a parameterized optimization problem [94].

The parameterized optimization problem is mathematically expressed in equation (4.1), in which the value function $v(\boldsymbol{t})$, a function of the parameters $\boldsymbol{t}$ is the minimum of the objective function $f(\boldsymbol{x}, \boldsymbol{t})$ with respect to the choice variables $\boldsymbol{x}$ at the given parameter setting. For the tow-steered composite technology performance problem the parameters are the planform variables and the choice variables are the global structural variables. The objective function is the structural weight with specified planform and global structural variables, and implicitly using optimal local structural variables. The value function is then the structural weight for the given planform variables, with optimal local and global structural variables.

$$v(\boldsymbol{t}) = \min_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{t}) \tag{4.1}$$

Whereas in a standard optimization problem the typical solution is a single point, a parameterized optimization problem is solved by finding the best point for each location in the parameter space. This reflects the idea that a system designer needs information about the structural performance for a variety of configurations in order to select the configuration, but is not concerned with the particular setting of structural variables that achieves that performance.

This parameterized optimization formulation should be used to guide the selection of new points. The surrogate model does not need to be globally accurate, but for every planform setting it must be accurate in the regions of the design space that help to predict the optimal tow paths.

4.1.3 Multistage Design of Experiments

The advantage of multistage experiments is that each stage can benefit from the information gathered in previous stages. This helps to focus the data collection in the most useful regions of the experimental space. Many different multistage procedures that have been developed to address a variety of experimental goals [95].

One of the simplest approaches is to add data with the goal of creating a globally accurate surrogate model. A globally accurate surrogate would completely characterize the system and allow any question about the system's behavior in the experimental region to be answered, even if that question was only thought of after the fact. Lam and Notz [96] developed a multistage experimental technique for this purpose, but in their experiments it was not much more effective than an equally sized space-filling design unless the response was noticeably nonstationary (violating a commonly used assumption about the correlations of a Gaussian process model). When combined with the finding of Loeppky, Sacks, and Welch [93] that inaccurate models improve slowly as data is added, this indicates that this approach would not work well for a tow-steered composite structure unless the inaccuracy is driven by nonstationary behavior. Although it would be ideal to have a globally accurate surrogate, it seems more practical to focus model accuracy in specific regions.

A review of the literature found no such multistage experiment specifically for parameterized optimization. However, one of the most common engineering tasks is optimization and a large number of multistage experiments have been developed for this purpose. This approach can be referred to as surrogate-based optimization. Although the motivation for these methods is finding the optimum more than creating a selectively accurate surrogate model, the latter is still obtained in pursuit of the former. These techniques have been surveyed by Forrester and Keane [97], and by Haftka, Villanueva, and Chadhuri [98]. One interesting method from Iuliano [99] takes a diversified approach, combining many previously developed criteria for selecting new points to cover any individual method's weaknesses.

A large number of surrogate-based optimization methods fall under the banner of Bayesian optimization. These create probabilistic models of the response and base their sampling decisions on it. Some early examples of these methods are the efficient global optimization [100] and sequential Kriging optimization [101] algorithms. Bayesian optimization techniques have proven very effective at using available information to hone in on

the optimum with a minimal number of analysis runs.

Using surrogate-based optimization to collect the data for a surrogate model of a tow-steered composite structure's weight is a sensible approach, but there are some important differences between optimization and parameterized optimization that need to be considered. In an optimization problem the solution is typically a point and as a result the region in which the surrogate needs to be accurate is very small compared to the overall search space. Even if there are multiple local minima, in engineering applications there are typically several attractive-looking basins rather than dozens or hundreds, resulting in several points where the surrogate must be accurate.

For parameterized optimization the solutions will fall on a hypersurface of the same dimensions as the parameters and consisting of the optimal points for each parameter setting. The surrogate model will need to be accurate over the entire hypersurface, and if there are multiple local minima there will be several such hypersurfaces along which the surrogate must be accurate. As a result it should be expected that, for a similarly complicated objective function, a much larger volume will need be accurately modeled and therefore more data will be needed to solve the parameterized optimization problem than the optimization problem.

A straightforward way to adapt surrogate-based optimization to the parameterized optimization problem would be to repeatedly apply it with different parameter settings. This leads to the question of how the parameter settings are chosen for each run of surrogate-based optimization, which could be answered by selecting another design of experiments for the parameters. However, this crossed approach of an outer parameter experiment and inner surrogate optimization experiment seems inefficient. In particular it seems overly restrictive for each sequence of runs within a surrogate optimization to use the same parameter settings instead of letting them vary more freely.

These observations about the difficulty of the tow-steered composite performance estimation problem and the lack of a multistage experimental design for parameterized optimization problems creates the following research question, which will be addressed in the rest of this

chapter:

RESEARCH QUESTION 2

What design of experiments should be used to sample data from the expensive physics-based analysis in order to efficiently make a surrogate model that can be used to accurately determine the weight of the optimal tow path structure for all possible planform settings?

Coming back to the Bayesian optimization methods, one of their greatest attributes is their adaptability. For a seemingly limitless number of applications they provide a unified approach: mathematically express the uncertainty in the solution and use that uncertainty to guide the search. Besides optimization, successful Bayesian methods have been developed for root finding [102], integration [103], and reliability analysis [104]. Methods have also been developed for multi-task optimization [105, 106], a type of problem with many similarities to parameterized optimization. In multi-task optimization there are environmental variables that play a similar role to the parameters, the difference is that the goal is to find the single design that performs best when averaged over the environmental settings, rather than separate best designs for each parameter setting.

Not only have Bayesian methods been developed for many different problems, they have also been successfully applied across a variety of disciplines, including machine learning [107], materials science [108], and aerospace design [109]. There is therefore plenty of room for optimism that this framework can be used to design a multistage experiment for a parameterized optimization problem.

4.2 Method Development

There is reason to think that Bayesian optimization can provide a good framework for developing a multistage design of experiments for making a surrogate model of tow-steered composite performance. The method will need to be specialized to parameterized optimization instead of regular optimization. This section will explain more of the details involved in

Bayesian optimization algorithms and how to adapt previous techniques to the problem at hand.

4.2.1 Bayesian and Gaussian Process Regression Models

Bayesian Regression

A Bayesian regression model is fundamental to Bayesian optimization and related methods. The key feature of these models is that their response is a probability distribution rather than a deterministic value. This allows them to express uncertainty in their predictions and enables multistage experimental strategies to avoid oversampling regions whose behavior, even if desirable, is already well understood.

A deterministic regression model is shown in equation (4.2), with predictions $\tilde{y}$ equal to some function f of the inputs $\boldsymbol{x}$ and the parameters $\boldsymbol{\theta}$. The form of the function is unspecified and could represent anything from a linear response surface to a multilayer neural network. The regression model is trained by adjusting its parameters to minimize some measure of the error, like the sum of the squared error.

$$\tilde{y} = f(\boldsymbol{x}; \boldsymbol{\theta}) \tag{4.2}$$

If the parameters are changed from deterministic to random variables, then the prediction also becomes a random variable. Its probability distribution is shown in equation (4.3) where the relationship between the inputs, parameters, and output is encoded in the conditional probability distribution $p(\tilde{y}|\boldsymbol{\theta}, \boldsymbol{x})$ and assumptions about the parameters are encoded in the prior distribution $p(\boldsymbol{\theta})$.

$$p(\tilde{y}|\boldsymbol{x}) = \int p(\tilde{y}|\boldsymbol{\theta}, \boldsymbol{x}) p(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} \tag{4.3}$$

Training of a Bayesian regression model consists of updating its parameter distribution based on the observations in accordance with Bayes' theorem [110], as shown in

equation (4.4a) with input data matrix X and observed response vector $\boldsymbol{y}$. The resulting prediction is shown in equation (4.4b). It is possible for the hyperparameters governing the prior distribution to themselves be modeled as random variables (with their distribution termed a hyperprior) in which case there are multiple layers of conditional probability. This can make the inference more difficult but the same math is used.

$$p(\boldsymbol{\theta}|\boldsymbol{y}, X) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta}, X)p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}|\boldsymbol{\theta}^*, X)p(\boldsymbol{\theta}^*)\,\mathrm{d}\boldsymbol{\theta}^*} \tag{4.4a}$$

$$p(\tilde{y}|\boldsymbol{x}, \boldsymbol{y}, X) = \int p(\tilde{y}|\boldsymbol{\theta}, \boldsymbol{x})p(\boldsymbol{\theta}|\boldsymbol{y}, X)\,\mathrm{d}\boldsymbol{\theta} \tag{4.4b}$$

The posterior probability is often difficult to calculate owing to the multidimensional integral in the denominator of equation (4.4a). In general, there is no nice analytic solution to this integral so sampling techniques like Hamiltonian Monte Carlo [111] are used to numerically approximate the posterior distribution. These techniques use the unnormalized posterior density in equation (4.5) which allows them to skip the troublesome integral.

$$p(\boldsymbol{\theta}|\boldsymbol{y}, X) \propto p(\boldsymbol{y}|\boldsymbol{\theta}, X)p(\boldsymbol{\theta}) \tag{4.5}$$

Gaussian Process Regression

Gaussian processes are recommended for use as a Bayesian regression models in the context of computer experiments [112], which makes them a prime candidate for modeling tow-steered composite performance. A Gaussian process [91] models the joint distribution of the response at a collection of input points as a multivariate normal random variable.

Because the Gaussian process uses the multivariate normal distribution it does have a nice analytic solution for equation (4.4b). As a result the inference step is both fast and accurate. Equations (4.6a) and (4.6b) show the Gaussian process prior and posterior

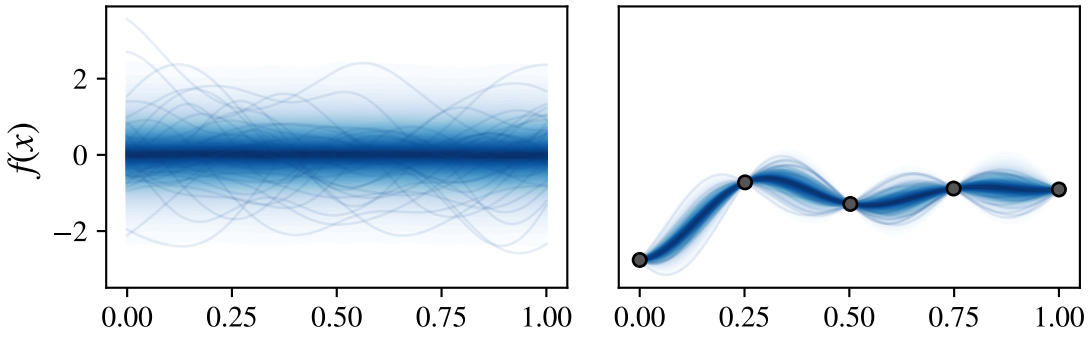

Figure 4.2: Gaussian process regression model. The shading and lines show the probability density and random functions sampled from the process, respectively, and the dots show the data. The left panel shows the prior and the right panel shows the posterior after conditioning on the data.

respectively, with mean function m and covariance function k.

$$\boldsymbol{y}|X \sim \mathcal{N}(\boldsymbol{m}(X), K(X,X)) \tag{4.6a}$$

$$\begin{aligned}\tilde{y}|\boldsymbol{x}, \boldsymbol{y}, X \sim \mathcal{N}(&m(\boldsymbol{x}) + \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{x}, X)K(X,X)^{-1}(\boldsymbol{y} - \boldsymbol{m}(X)),\\ &k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{x}, X)K(X,X)^{-1}\boldsymbol{k}(X, \boldsymbol{x}))\end{aligned} \tag{4.6b}$$

An example of Gaussian process inference is shown in figure 4.2. The prior admits a wide variety of possibilities but after conditioning on the data the output distribution is much narrower, especially around the training data. A range of nonlinear behaviors are possible in the gaps between. Note that models with simple assumed functional forms, like a quadratic polynomial, would struggle to fit the data in this example without explaining a lot of the variation as noise. Although the example's data is not necessarily representative of tow-steered composite behavior, the modeling capability it illustrates is applicable.

Gaussian processes provide great flexibility to model different system behaviors through selection of their mean and covariance functions. There are many possible forms of covariance, or kernel, function [113]. In engineering applications they typically embody the idea

that nearby points in the input space have similar responses but other behaviors, like periodic correlation, are possible. There are also usually several hyperparameters associated with covariance functions; in the case of spatial correlation there will be a parameter controlling how quickly the response can vary. The hyperparameters of the mean and covariance functions are usually not known beforehand so must be accounted for via maximum likelihood estimation, Monte Carlo sampling, or some other method [114].

Equation (4.7) shows perhaps the most commonly used covariance function, the squared exponential. In this parameterization of the squared exponential the magnitude hyperparameter σ controls the scale of the response and the sensitivity hyperparameters θ_i control how quickly the response can vary across each dimension.

$$K(\boldsymbol{x}, \boldsymbol{x}^*) = \sigma^2 \exp(-\sum_{i=1}^{d} \theta_i (x_i - x_i^*)^2) \tag{4.7}$$

Figure 4.3 illustrates the variety of behaviors enabled by covariance function selection. Each curve shows a single random function drawn from a Gaussian process with the specified characteristics. The left side shows the difference that results from changing the sensitivity hyperparameter. As the sensitivity increases the function changes more rapidly. For tow-steered composite performance, it is possible that the aspect ratio has a low sensitivity associated with it, resulting in a gradual monotonic weight trend while a tow path coefficient has a high sensitivity, resulting in more rapid variation.

The right side of figure 4.3 shows the difference due to the form of the covariance. A couple of Matérn covariances are shown alongside the squared exponential. The Matérn covariances generalize the squared exponential and permit rougher behavior, especially the Matérn ½. Although this roughness makes it more difficult to extrapolate the response in the areas between training points, it is possible that it would be a better representation of the tow-steered composite weight analysis. The response might look rough because of:

- discrete changes in structural topology as ribs appear or disappear with changes in the planform's span

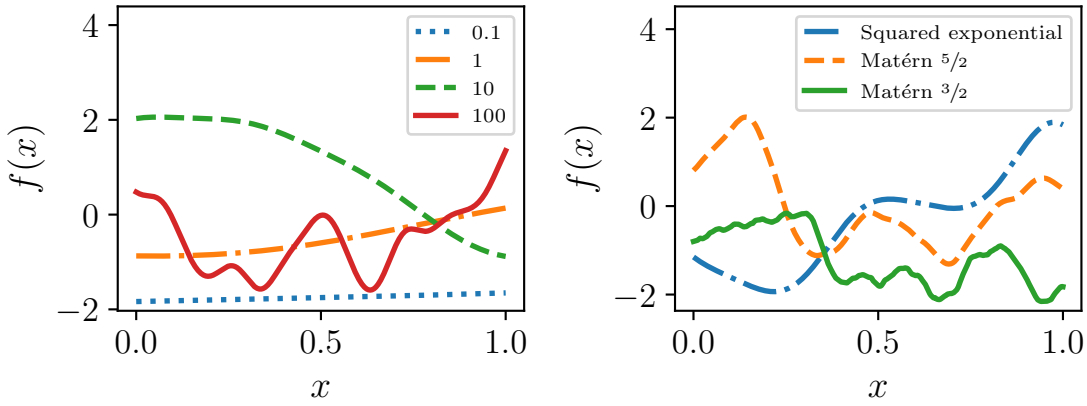

Figure 4.3: Functions sampled from Gaussian processes with different covariances. The left side shows functions drawn from Gaussian processes with a squared exponential covariance function but different sensitivities. The right side shows functions drawn from Gaussian processes with different covariance function forms.

- discrete changes in the mesh of the structural finite element model or aerodynamic vortex lattice model
- numerical tolerances used in geometry construction
- numerical tolerances for convergence of the structural analysis and design loop

As an alternative to using a rough covariance function, it is also possible to account for these features by modeling the tow-steered composite response as noisy observations of a smooth underlying function.

The many possible mean and covariance structures can be compared to find which combination (and hyperparameter setting) provides the best explanation of the observed data.

Model Validation

Regardless of the covariance form and hyperparameters that are ultimately selected, the regression model should be checked to validate the assumptions. Leave-one-out cross-validation, in which the model predicts each training point while holding it out of the

training set, has been recommended for this purpose [100] and is particularly efficient in the case of Gaussian process models. The leave-one-out prediction accuracy should give an idea of the model's performance at unevaluated points, and if the data comes from a space-filling design of experiments there is reason to think that the leave-one-out accuracy is slightly pessimistic compared to the accuracy at unevaluated points. The results of the leave-one-out procedure can be judged graphically using the actual vs. predicted plot and normal quantile plot.

An example of an actual vs. predicted plot is displayed in the left panel of figure 4.4. It consists of the actual-equals-predicted line, dots representing the data's actual and predicted values, and intervals showing the model's uncertainty for each prediction. A common choice for displaying the prediction interval is to use three standard deviation limits.

If the predictions group tightly along the actual-equals-prediction line, then the surrogate is very accurate and the followup experimental stages are unnecessary. If the predictions roughly follow the actual-equals-predicted line, then the surrogate has learned trends in the data and a multistage experiment should be used to improve accuracy. If the predictions do not align with the actual-equals-prediction line, then the surrogate model is not working well and the experimenter will need to investigate why that is.

The prediction intervals can indicate if the model is underconfident or overconfident. An underconfident model's actual vs. predicted plot will show the predictions closely following the actual-equals-prediction line, but the prediction intervals will be wide. An overconfident model will have predictions that do not follow the actual-equals-predicted line very well, but will have narrow prediction intervals that do not extend to the actual-equals-predicted line.

The actual vs. predicted plot can help identify outliers in the data. If the prediction interval for any of the points does not overlap or at least come close to the actual-equals-prediction line, then that point is strongly inconsistent with the rest of the data and therefore an outlier. The details of the corresponding simulation execution should be investigated to determine if there was an error that resulted in invalid data. As a hypothetical, the

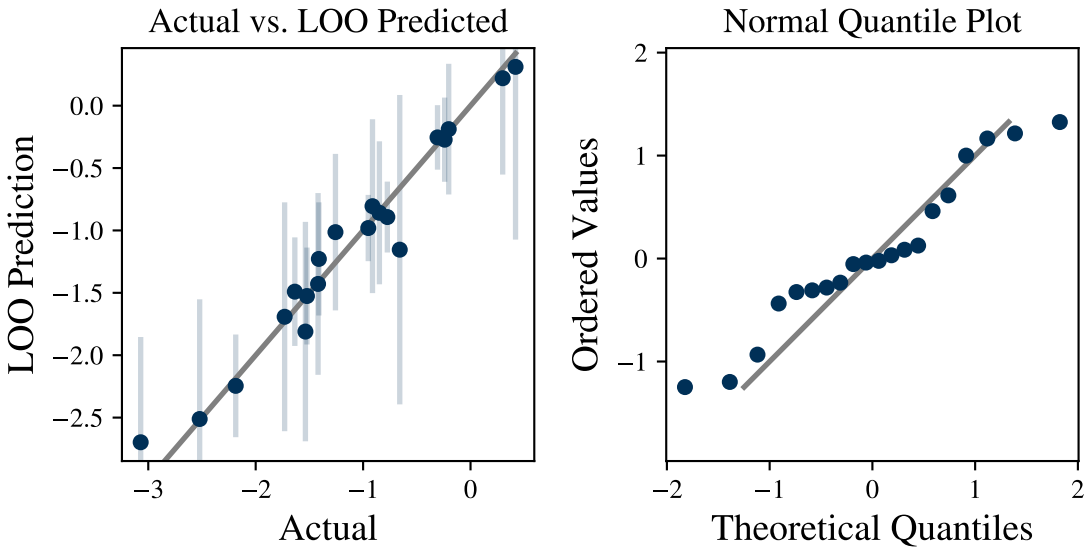

Figure 4.4: Diagnostic plots for checking surrogate model adequacy. The left panel shows an actual vs. predicted plot, with actual-equals-predicted line and prediction intervals for each data point. The right panel shows a normal quantile plot which indicates how well the residuals fit a normal distribution.

tow-steered composite analysis could have created a poor mesh that resulted in a stress concentration and spurious high weight prediction. If investigation finds that the data is invalid, it should be removed from the training set. If the data is valid then the statistical model is misspecified and will need to be modified so that it represents the data better.

The actual vs. predicted plot (or a residual by predicted plot) can also help determine if the output data should be transformed. For instance, if the size of errors increases with the magnitude of the response, that is a sign that the data should be log-transformed.

An example of a normal quantile plot is shown on the right side of figure 4.4. The purpose of this plot is to show how well the surrogate's prediction variance reflects its actual accuracy. It does this by showing the sorted residuals, normalized by the prediction standard deviation, against the theoretical quantiles of a standard normal distribution. For a good model the data should line up on the empirical-equals-theoretical line, indicating that the surrogate model has a good understanding of its accuracy. If the data is more vertical

that indicates that the surrogate is too confident in its predictions and if the data is more horizontal that indicates that the model is not confident enough. This is important for a multistage sampling procedure because the uncertainty in the model's predictions helps to choose the next point added. An underconfident model might lead to wastefully sampling regions of the input space that are already well understood, and an overconfident model might lead to neglecting to sample regions that actually have a significant error. The normal quantile plot can also be use to identify outliers by looking for individual points far from the line.

4.2.2 Bayesian Optimization

The collection of Bayesian optimization and related methods follow a consistent overall process. Assuming the input factors and their ranges have been selected, the first step is to run an initial design of experiments and train the surrogate model, which should then be checked to ensure that it is working correctly (this does not mean it must be perfectly accurate, if that were the case there would be no need to add followup samples). Since Bayesian optimization methods are typically applied to difficult functions it makes sense to choose a space-filling design. It also makes sense to choose a standard size for the initial stage: though there is logic to sampling a small initial stage and quickly proceeding to the adaptive stages, there is a risk that without enough data, the hyperparameter estimates and the surrogate model will be poor and the adaptive sampling will be ineffective.

After the initial stage, the adaptive stages take advantage of the current information when picking which data to add. The input space is searched to find the maximum of what is called the acquisition function. The job of the acquisition function is to quantify how beneficial it would be to add a candidate point to the training data. It is through different acquisition functions that Bayesian methods have been adapted to different problem types. Even for a single problem type there are often alternative acquisition functions to choose from.

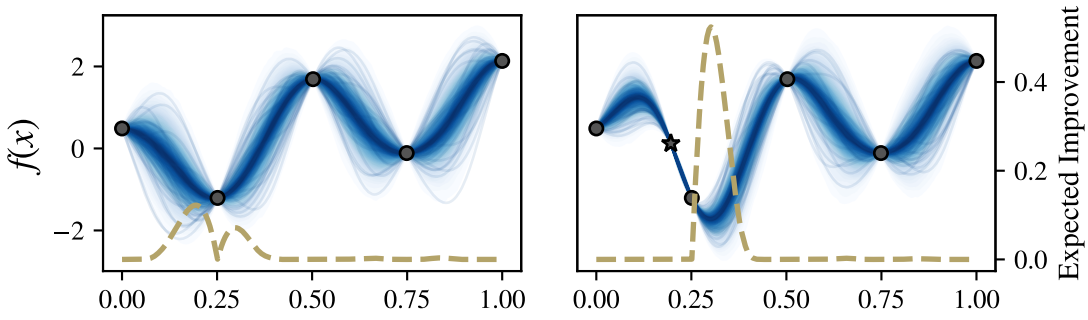

Figure 4.5: An iteration of Bayesian optimization using the expected improvement acquisition function. The gray points show the observations, the blue shading and lines show the probability density and samples from the Bayesian model, respectively, while the dashed gold lines show the expected improvement. The left panel shows the situation after the initial data collection. The right panel shows the result of adding the point with the highest expected improvement, indicated by the star.

Figure 4.5 shows an example iteration of Bayesian optimization using the expected improvement [100] acquisition function. Since this acquisition function was developed for optimization problems it is expected that it is greater in regions that appear more likely to contain the optimum. In this example, even though the added point does not provide direct improvement, it still provides a lot of information about the location of the solution.

Data can keep being added in sequential experimental stages until some exit condition is achieved. This might be when the maximum of the acquisition function is small, indicating that there is very little to gain from additional data, or some other problem-specific criterion.

This procedure was presented as the following steps in the article presenting the sequential Kriging optimization (SKO) algorithm [101] (Kriging refers to Gaussian process regression):

1. Build an initial Kriging meta-model of the objective function.
2. Use cross-validation to ensure that the Kriging prediction and measure of uncertainty are satisfactory.
3. Find the location that maximized the expected improvement (EI) function. If the maximal EI is sufficiently small, stop.

4. Add an evaluation at the location where the EI is maximized. Update the Kriging meta-model using the new data point. Go to step 3.

This procedure could be used to solve the parameterized optimization problem associated with tow-steered composite performance. All that needs to change is that the expected improvement acquisition function must be swapped for a more suitable acquisition function.

4.2.3 Acquisition Functions

The acquisition function is how the preference about which point to add is mathematically expressed. For tow-steered composite performance evaluation the acquisition function should favor points that help to the parameterized optimization problem. Although no such acquisition functions was found in literature, there are many for ordinary optimization and it might be possible to modify one of those to embody the different problem type.

There are many acquisition functions that have been developed for Bayesian optimization. While some are optimal in a statistical or information theory sense, most are intuitive heuristics that have nonetheless proven effective. There is a range of complexity and computational expense to the acquisition functions: while Bayesian optimization presumes an expensive objective function, the magnitude of this expense matters and the more expensive acquisition functions might not be worthwhile for relatively cheaper objective functions.

Notationally, because it is possible for function evaluations to be noisy, a distinction is drawn between the observation y and the unobserved underlying signal f. Since the signal is unknown it has epistemic uncertainty that is represented probabilistically with mean μ and standard deviation σ. Future observations are also probabilistic, but past observations are deterministic.

One of the simplest acquisition functions is the probability of improvement described by Mockus, Mockus, and Mockus [115]. Shown in equation (4.8a), it measures the probability that the proposed point has a lower objective function value than the best observation so far.

For normally distributed predictions, like in a Gaussian process model, the probability has a nice analytical form show in equation (4.8b) where Φ is the normal cumulative distribution function. Since probability of improvement compares the surrogate's prediction to the best observation so far, it can only be used with noiseless objective functions; otherwise, there is a chance that the current best point would only appear so due to a very favorable noise realization. An issue with the probability of improvement is that the point it can favor points with a high chance of providing negligible improvement. A term can be added to require a specific amount of improvement, but then a heuristic for setting that term must be added as well.

$$\mathbb{PI}(\boldsymbol{x}) = p\left(f(\boldsymbol{x}) \leq \min_i y_i\right) \tag{4.8a}$$

$$\mathbb{PI}(\boldsymbol{x}) = \Phi\left(\frac{\min_i y_i - \mu(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\right) \tag{4.8b}$$

The expected improvement criterion shown in equation (4.9) is the acquisition function used in the efficient global optimization algorithm [100]. It also has a nice analytic form in the case that the prediction is normally distributed. Like probability of improvement it is simple and intuitive; it measures the average amount that the candidate point would improve the best observation. Since it rewards points that provide more improvement, it is more prone to exploration than exploitation compared to probability of improvement.

$$\mathbb{EI}(\boldsymbol{x}) = \mathbb{E}\left[\max\left(0, \min_i y_i - f(\boldsymbol{x})\right)\right] \tag{4.9}$$

Similarly to probability of improvement, expected improvement requires assuming that the observations from the objective function are noise-free. This assumption is not used in sequential Kriging optimization [101], which uses the modified form of expected improvement shown in equation (4.10) that allows for the observations to have noise variance

${\sigma_\epsilon}^2$. The factor k used in the calculation of $\boldsymbol{x}^{**}$ is set to a positive value to make the reference minimum conservative.

$$\mathbb{EI}(\boldsymbol{x}) = \mathbb{E}[\max(0, f(\boldsymbol{x}^{**}) - f(\boldsymbol{x}))] \cdot \left(1 - \frac{\sigma_\epsilon}{\sqrt{\sigma^2(\boldsymbol{x}) + {\sigma_\epsilon}^2}}\right) \tag{4.10}$$

$$\text{with} \quad \boldsymbol{x}^{**} = \underset{\boldsymbol{x} \in \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n}{\operatorname{argmin}} \left[\mu(\boldsymbol{x}) + k\sigma(\boldsymbol{x})\right]$$

Lower confidence bound [116] (or upper confidence bound, if the objective is maximized) uses an optimistic quantile from the surrogate's predictive distribution. Unlike the other acquisition functions, lower confidence bound directly corresponds to the function being optimized. It is invariably used in conjunction with normal predictive distributions, as seen in equation (4.11), because the search's optimism can easily be expressed through the coefficient k on the prediction standard deviation. This equation has the same form as the one used to find the reference minimum in the noisy version of expected improvement, but in this case the factor is set to a negative value to make the prediction optimistic rather than conservative.

$$LCB(\boldsymbol{x}) = \mu(\boldsymbol{x}) + k\sigma(\boldsymbol{x}) \tag{4.11}$$

The knowledge gradient [117], shown in equation (4.12), comes from a similar intuition as expected improvement but generalizes it and is a bit more complicated. The mean prediction of the current surrogate μ_n is deterministic and is minimized to serve as a reference value. In contrast, the next-stage mean μ_{n+1} is random because it depends on a future observation at the candidate point, which is modeled as a random variable. Each possible realization of the observation corresponds to a realization of the next-stage surrogate's mean, and each of those in turn has its own minimum. The knowledge gradient is calculated by taking the expected value of the next-stage minimum and subtracting that from

the reference value. The overall effect is to measure the expected decrease in the minimum of the surrogate's mean prediction due to adding the candidate point. To clarify the difference between expected improvement and knowledge gradient, expected improvement calculates improvement at the candidate point relative to the best observation so far and knowledge gradient calculates improvement in the mean prediction's minimum relative to the current minimum. A couple of distinguishing characteristics of knowledge gradient are that it inherently accounts for noisy objective functions without needing a tuning parameter and that it trusts the probabilistic model when it makes confident predictions about unsampled locations.

$$\mathbb{KG}(\boldsymbol{x}) = \min_{\boldsymbol{x}^*} \mu_n(\boldsymbol{x}^*) - \mathbb{E}\left[\min_{\boldsymbol{x}^{**}} \mu_{n+1}(\boldsymbol{x}^{**}) | \boldsymbol{x}_{n+1} = \boldsymbol{x}\right] \tag{4.12}$$

A difficulty of using knowledge gradient is that each evaluation requires solving an optimization problem. Since knowledge gradient must itself be optimized to choose the next point, this creates a bilevel optimization problem. While knowledge gradient was originally developed for discrete choice problems, many problems of interest including the one in this work are continuous choice problems. In order to avoid a challenging continuous bilevel optimization problem, the knowledge gradient for continuous parameters [118] was developed, shown in equation (4.13). This uses the heuristic of limiting the inner search to a discrete set of points in order to simplify the computation. The set of points considered in the inner optimization consist of the previously sampled points as well as the currently proposed point.

$$\mathbb{KGCP}(\boldsymbol{x}) = \min_{\boldsymbol{x}^* \in \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{x}\}} \mu_n(\boldsymbol{x}^*) - \mathbb{E}\left[\min_{\boldsymbol{x}^{**} \in \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{x}\}} \mu_{n+1}(\boldsymbol{x}^{**}) | \boldsymbol{x}_{n+1} = \boldsymbol{x}\right] \tag{4.13}$$

Predictive entropy search [119] is perhaps the most theoretically direct acquisition function for Bayesian optimization, and should beat the other acquisition functions when

measured by number of objective function calls. However it is also more expensive to evaluate, so is most appropriate for very expensive objective functions. This acquisition function is based on the fact that the Bayesian regression model implicitly defines a probability distribution for the location of the optimum, and the uncertainty of this distribution can be quantified by its entropy. There is no nice analytic form for this distribution so it must be estimated, for instance, with Monte Carlo sampling and kernel density estimation. Predictive entropy search then chooses the sample that reduces the entropy the most.

Since these acquisition functions were developed for optimization rather than parameterized optimization, they are only able to assess the relative value of adding points at different locations in the choice variable space. An acquisition function for parameterized optimization will need to build on this so that it can also appropriately judge the value of different locations in the parameter space. The intuition is that the acquisition function should prefer points that help narrow down the optimal choice variables over as wide a region of the parameter space as possible.

The idea of covering the parameter space invokes the concept of integration, as in Bayesian multi-task optimization [106]. Integrating an acquisition function for Bayesian optimization over the parameter space should yield an acquisition function for parameterized optimization.

The acquisition functions surveyed in this section were evaluated to determine their suitability for use in a Bayesian parameterized optimization method. There was a strong motivation to use expected improvement since it is such a standard choice for Bayesian optimization. However, it compares the proposed sample to the best so far, and in order for the comparison to be fair the proposed point and the comparison point would need to have the same parameter setting. This is a problem because only a discrete set of points in the parameter space will even have a sample, much less one that is representative of the "best so far"

The modified expected improvement from equation (4.10) is better suited because it uses

the surrogate to calculate the comparison value; it could be further modified by projecting the existing data points so that they sit at the same location in parameter space as the proposed point. After doing that it would strongly resemble the knowledge gradient for continuous parameters but with the added disadvantage of having the factor k that must be subjectively tuned. This observation lends support to the use of the knowledge gradient for continuous parameters.

Probability of improvement is limited in that it does not account for the magnitude of improvement that a sample can provide, and the way to remedy this is to add a term that must be tuned ad hoc. Similarly, lower confidence bound depends on a factor that must be tuned ad hoc.

Predictive entropy search would probably solve the problem with a minimal number of additional cases. However there was a concern that it might be computationally expensive and difficult to implement. For optimization, it already needs to sample enough optima from realizations of the Bayesian model to construct an empirical distribution. For parameterized optimization this would need to be repeated at many locations in parameter space in order to perform the integration.

4.2.4 Integrated Knowledge Gradient

Therefore, the knowledge gradient for continuous parameters was selected as the acquisition function to adapt for use in a Bayesian parameterized optimization method, since it does not require any ad hoc tuning and it makes comparisons between values of the surrogate rather than the data directly. If it were to be applied to a parameterized optimization problem, the parameters could be handled like in the form shown in equation (4.14) where a $\boldsymbol{t}'$ is chosen to be the parameter vector at which the criterion is evaluated and $\boldsymbol{t}$ is the parameter vector of the candidate point. For a given evaluation parameter vector $\boldsymbol{t}'$ it would make a lot of sense to select a candidate point with the same parameters, since that would provide the most direct information about that point in parameter space, but there is no requirement that

$\boldsymbol{t} = \boldsymbol{t}'$.

$$\mathbb{KGCP}(\boldsymbol{t}, \boldsymbol{x}; \boldsymbol{t}') = \min_{\boldsymbol{x}^* \in \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{x}\}} \mu_n(\boldsymbol{t}', \boldsymbol{x}^*) - \mathbb{E}\left[\min_{\boldsymbol{x}^{**} \in \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{x}\}} \mu_{n+1}(\boldsymbol{t}', \boldsymbol{x}^{**}) \;\middle|\; \{\boldsymbol{t}_{n+1}, \boldsymbol{x}_{n+1}\} = \{\boldsymbol{t}, \boldsymbol{x}\} \right] \quad (4.14)$$

However, this form is not satisfactory because it only considers a single point in the parameter space, whereas the parameterized optimization problem is concerned with the entire parameter space. In order to account for the entire parameter space when selecting a new data point it is proposed to integrate over $\boldsymbol{t}'$, resulting in the integrated knowledge gradient ($\mathbb{IKG}$) shown in equation (4.15).

$$\mathbb{IKG}(\boldsymbol{t}, \boldsymbol{x}) = \int_{\boldsymbol{t}'} \cdots \int \left(\min_{\boldsymbol{x}^* \in \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{x}\}} \mu_n(\{\boldsymbol{t}', \boldsymbol{x}^*\}) - \mathbb{E}\left[\min_{\boldsymbol{x}^{**} \in \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{x}\}} \mu_{n+1}(\{\boldsymbol{t}', \boldsymbol{x}^{**}\}) \;\middle|\; \{\boldsymbol{t}_{n+1}, \boldsymbol{x}_{n+1}\} = \{\boldsymbol{t}, \boldsymbol{x}\} \right] \right) \mathrm{d}\boldsymbol{t}' \quad (4.15)$$

Although this form clearly shows how to obtain $\mathbb{IKG}$ from $\mathbb{KGCP}$, it is not convenient to work with. For instance, it might appear that evaluating the next-stage surrogate mean μ_{n+1} at each candidate point would require drawing values from the current surrogate's posterior and then recalculating the conditional probability distribution. This would be very costly since each $\mathbb{IKG}$ evaluation would require a matrix factorization, but fortunately there is a much easier way. Wu and Frazier [120] show how to express μ_{n+1} in terms of the current surrogate's mean μ_n and covariance $\tilde{\sigma}_n$, which can be evaluated at arbitrary points cheaply, and a standard normal random variate z. Another helpful change is to express the integral as an equivalent expected value by interpreting the integration variable $\boldsymbol{t}'$ as a uniformly distributed random variable and conditioning the inner expectation on this value. These two

changes transform the $\mathbb{IKG}$ into equation (4.16). Note that $\boldsymbol{t}$ and $\boldsymbol{x}$ refer to the candidate sample location, $\boldsymbol{x}^*$ and $\boldsymbol{x}^{**}$ refer to the minima of the current and next-stage surrogates, and $\boldsymbol{t}'$ is a dummy parameter vector.

$$\mathbb{IKG}(\boldsymbol{t},\boldsymbol{x}) = \mathbb{E}\Bigg[\min_{\boldsymbol{x}^*\in\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{x}\}} \mu_n(\{\boldsymbol{t}',\boldsymbol{x}^*\}) - \mathbb{E}_z\left[\min_{\boldsymbol{x}^{**}\in\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{x}\}} \left(\mu_n(\{\boldsymbol{t}',\boldsymbol{x}^{**}\}) + z\tilde{\sigma}_n(\{\boldsymbol{t}',\boldsymbol{x}^{**}\},\{\boldsymbol{t},\boldsymbol{x}\})\right) \;\middle|\; \boldsymbol{t}'\right]\Bigg] \tag{4.16}$$

Finally, equation (4.16) can be approximated by Monte Carlo sampling. Due to the linearity of expectation and the independence of $\boldsymbol{t}'$ and z, the inner expectation can be expanded to include the first term and then combined with the outer expectation so that there is only one level of sampling with m draws, resulting in the approximation shown in equation (4.17).

$$\mathbb{IKG}(\boldsymbol{t},\boldsymbol{x}) \approx \frac{1}{m}\sum_{i=1}^{m}\Big(\min_{\boldsymbol{x}^*\in\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{x}\}} \mu_n(\boldsymbol{t}_i',\boldsymbol{x}^*) - \min_{\boldsymbol{x}^{**}\in\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n,\boldsymbol{x}\}} \left(\mu_n(\boldsymbol{t}_i',\boldsymbol{x}^{**}) + z_i\tilde{\sigma}_n(\{\boldsymbol{t}_i',\boldsymbol{x}^{**}\},\{\boldsymbol{t},\boldsymbol{x}\})\right)\Big) \tag{4.17}$$

4.2.5 Method Description

Due to nonlinearity and high dimensionality it is difficult to construct an accurate surrogate model of tow-steered composite wingbox weight. Analyzing the system is computationally expensive, so data collection must be carefully planned so that as much useful information as possible is gained from each case. Bayesian optimization methods have been applied to a wide variety of problem types across many fields of study, and have proven to be effective at learning as much as possible from each selected case.

Therefore, a method based on adapting Bayesian optimization to a parameterized optimization problem has been developed. To guide the selection of new points, the integrated knowledge gradient acquisition function has been developed. The overall algorithm, called Bayesian parameterized optimization, is shown in algorithm 2. It is mainly based on the procedure from sequential Kriging optimization [101].

It is recommended that the Monte Carlo samples are only drawn once per iteration and held constant while optimizing $\mathbb{IKG}$ so that the optimization is deterministic. Even if there is sampling bias within an iteration, over the course of many iterations those kinds of effects should average out.

Like in other Bayesian methods, and as will be seen in while experimentally testing this method, $\mathbb{IKG}$ has multiple local maxima. Therefore a global optimization strategy should be used to maximize $\mathbb{IKG}$ and find the next data point to collect. Options include differential evolution, particle swarm, or gradient-based algorithms with multiple starts. As noted by Frazier, Powell, and Dayanik [117], it is not terribly important to find the point that maximizes the acquisition function, as long as good solutions can be found.

Algorithm 2 Bayesian Parameterized Optimization

Require: Parameterized optimization problem with parameters $\boldsymbol{t}$, choice variables $\boldsymbol{x}$, objective function $f(\boldsymbol{t}, \boldsymbol{x})$, and bounds on $\boldsymbol{t}$ and $\boldsymbol{x}$; number of multistage cases n, and number of samples m

Run initial design of experiments
Validate the initial Bayesian regression model, optionally comparing and selecting from model alternatives
for $j = 1, n$ **do**
    Generate m samples $(\boldsymbol{t}'_i, z_i)$
    Using a global optimization strategy, find $(\boldsymbol{t}^\dagger, \boldsymbol{x}^\dagger)$ that maximizes $\mathbb{IKG}$
    if The maximal $\mathbb{IKG}$ is sufficiently small **then**
        break
    end if
    Evaluate $f(\boldsymbol{t}^\dagger, \boldsymbol{x}^\dagger)$
    Update the surrogate model with the new data point
end for

The following hypothesis statement expresses the belief that this method will succeed in its intended purpose:

HYPOTHESIS 2

If Bayesian parameterized optimization is used to drive data collection for the tow-steered composite design problem, then the data will be collected efficiently and the resulting surrogate model will be able to recover accurate solutions of the parameterized optimization problem, because Bayesian optimization methods are able to focus the improvement of surrogate model accuracy in regions of the design space that are likely to contain solutions.

An important caveat to this statement is that any experimental test of this hypothesis is jointly a test of the specific implementation of $\mathbb{IKG}$ used in the experiment. Since it is impractical to directly evaluate the integrated knowledge gradient as shown in equation (4.16), it must be approximated through sampling and other heuristics.

The remaining sections of this chapter describe efforts to validate the method and confirm or reject hypothesis 2. Section 4.3 presents an example that helps illustrate how the method works. Section 4.4 explains an experiment on artificial test functions that sheds some light on how characteristics of the objective function affect the success of this method. Finally, section 4.5 applies the method to the physics-based testbed from section 3.3 in order to determine if the method succeeds on the intended problem.

4.3 Branin Function Proof of Concept

In order to illustrate how Bayesian parameterized optimization works, and to begin to assess its effectiveness, the method was applied to the Branin function. The Branin function [121] is shown in equation (4.18) along with typical ranges for the input variables and is a standard test problem for global optimization algorithms. It is a contrived function so any results in this section do not necessarily translate to the intended application, but there are a couple advantages to using it at this stage. First, it is inexpensive to compute so the method can be run all the way through in a short time. This allows for iteratively adjusting the method and getting rapid feedback on the impact of the adjustments. Second, it is two-dimensional so

quantities of interest can be plotted over the entire space to get a complete picture of the method's behavior.

$$f(x_1, x_2) = (x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})\cos(x_1) + 10 \tag{4.18}$$

$$x_1 \in [-5, 10],\ x_2 \in [0, 15]$$

To adapt the Branin function into a parameterized optimization problem's objective function the original input variables x_1 and x_2 are assigned to the choice variable x and parameter t, respectively. This transposes the typical order of the variables, but in the author's opinion this creates a more interesting example. A contour plot of the objective function and a plot of the value function are shown in figure 4.6. It has three separate local minima that lie along a parabolic valley. Two of these local minima are on the left side and cause the optimization problems at these parameter settings to have multiple local minima. Because of these multiple local minima there are two locations where the optimal choice variable setting makes a discontinuous jump into another basin. One of these jumps creates a sharp cusp in the value function, which illustrates how a value function can have discontinuous derivatives even if the original function is smooth everywhere.

Walk-Through

Figure 4.7 steps through the progression of Bayesian parameterized optimization over the first few iterations. Following algorithm 2, the first step is to run the initial design of experiments, which was chosen to be a maximin Latin hypercube with 16 cases.

The chosen form for the surrogate model was a Gaussian process with constant mean, a squared exponential covariance, and an assumption of no noise in the observations. The hyperparameters of this model were fit at each iteration using maximum likelihood estimation. Although the second step in algorithm 2 calls for validating the model, this was

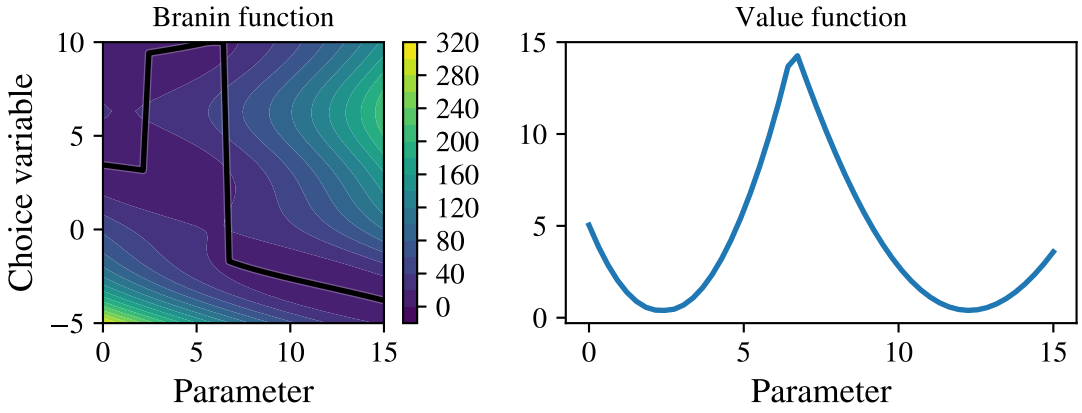

Figure 4.6: Branin function used in a parameterized optimization problem. The left panel shows the contours of the Branin function and traces the trajectory of optimal choice variables across parameter settings. The right panel shows the value function which is the optimal value for each parameter setting.

not strictly followed as experience already shows that this model form works well for this function. The first two panels in the top row of figure 4.7 show the model's prediction using the initial data set. The left shows the predicted objective function, which coarsely recreates parabolic valley from the original function though the details are different. The middle panel shows the model's value function which is somewhat accurate for most parameters but noticeably more inaccurate at high parameter settings.

Next, $\mathbb{IKG}$ was approximated throughout the input space using a total of 10,000 Monte Carlo samples. These samples used only 100 unique values of the parameter, which were resampled and combined with antithetic samples of the standard normal variate. The next section on sampling details explains the advantages of using a smaller number of unique parameter values and antithetic standard normal variates.

The contours of $\mathbb{IKG}$, shown in the right panels of figure 4.7, reveal complicated behavior. The response has noticeable saddle points at each data location with filament-like patterns running between each point and its neighbors. The filaments aligned with the valley tend to contain local maxima while the filaments that cut across the valley tend to contain local

minima. The intuitive interpretation of this behavior is that cross sections of the valley with a point on either side have less uncertainty than cross sections of the valley with no nearby points. There is also a distinct aversion to sampling from the upper-right or lower-left; both are maxima of the objective function surrogate and unlikely to be near solutions of the parameterized optimization problem. This plot also shows why a global optimization strategy should be used to maximize $\mathbb{IKG}$, since the values of the local maxima vary greatly. The maximum is indicated by a star and is added to the data before starting the next iteration.

For the first iteration, the chosen point is at a higher parameter setting and at a choice variable setting that is near the minimum. The plot in the middle of the second row shows that adding this point greatly reduces the error in the surrogate value function that occurred at higher parameter values. The next iteration proceeds just like the previous one, with newly drawn Monte Carlo samples, and selects a point at the middle of the parameter space and near the upper optima in the choice variable space. After adding just these two points, the objective function surrogate has become more accurate and the value function surrogate has greatly improved.

Monte Carlo Sampling Details

During this stage of developing Bayesian parameterized optimization, several sampling techniques were found to provide an advantage over simple random sampling.

First, it was observed that simple random sampling could produce the spurious result of the $\mathbb{IKG}$ approximation being negative. As explained by Scott, Frazier, and Powell [118] the knowledge gradient is nonnegative, so the integral of a nonnegative quantity over the parameter variables must also be nonnegative. Looking at the approximation formula (4.17), an individual sample is negative when the next iteration's surrogate has a higher minimum than the current surrogate, which can happen when the product of the covariance and the normal variate is positive. This a perfectly legitimate outcome, and corresponds to the real possibility that adding another point to the data set can increase the surrogate's minimum.

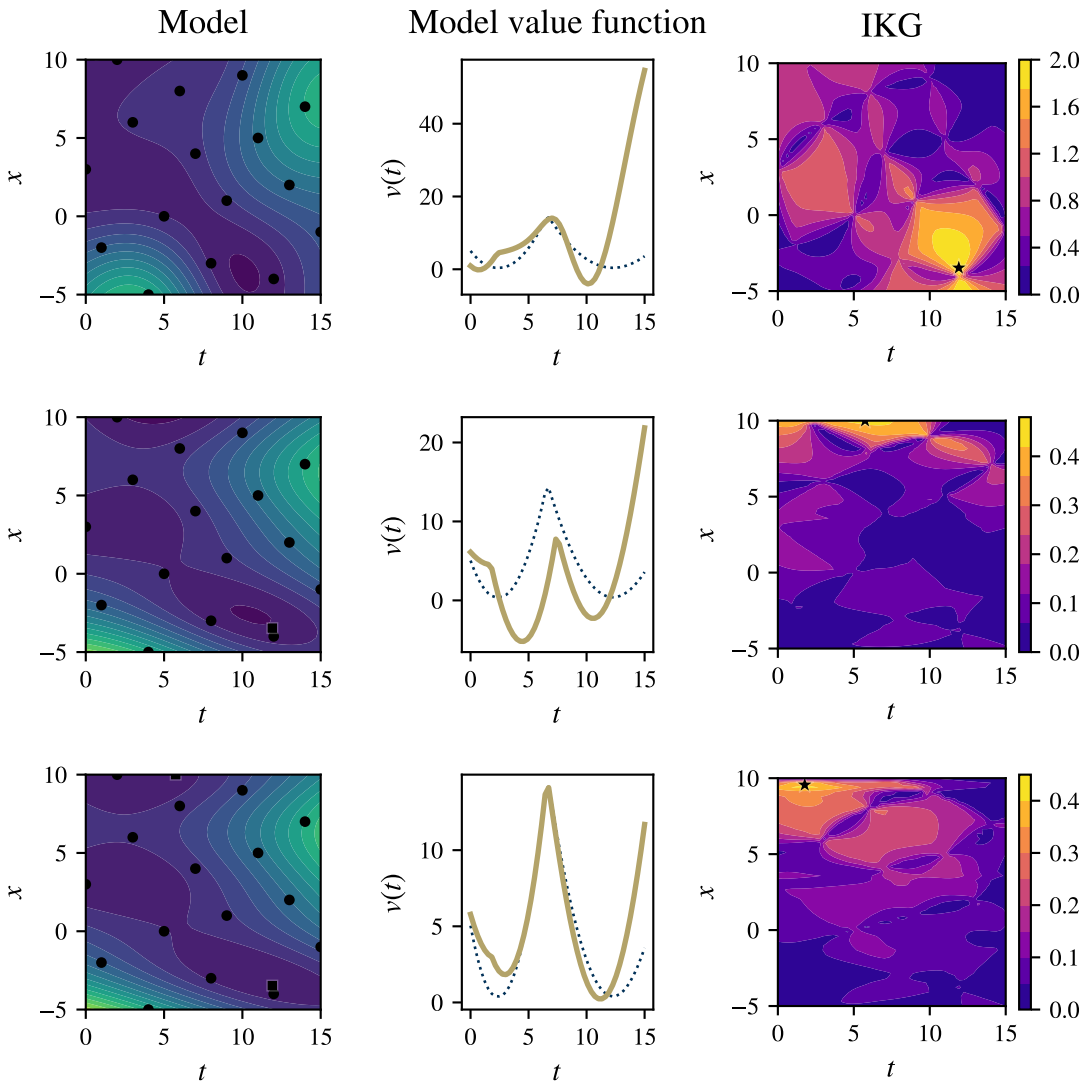

Figure 4.7: Bayesian parameterized optimization of the Branin function. Each row shows an iteration of the method starting immediately after the initial experimental stage. The left column shows the contours of the iteration's surrogate model, as well as the initial and sequentially added data points, represented with circles and squares respectively. The middle column shows the iteration's value function surrogate over a dotted line representing the true value function. The panels in the right column show contours of the $\mathbb{IKG}$ acquisition function and indicate the location of the maximum with a star. Note that the color scale for the contour levels of $\mathbb{IKG}$ changes in each row.

However the chance that the surrogate's minimum increases is balanced by at least as much probability that the surrogate's minimum decreases. With a large enough set of samples, each negative case should be balanced by an equal or greater magnitude positive case.

In order to remedy this behavior, antithetic sampling [122] of the normal variate z was investigated. Antithetic sampling is a variance reduction technique that creates opposing pairs of samples, so if one sample is $\{t'_i, z_i\}$ then another sample will be $\{t'_i, -z_i\}$. This actively enforces the condition that each sample where the minimum increases is balanced by a sample where the minimum decreases by at least as much, rather than passively hoping that it will balance out due to the law of large numbers.

The effectiveness of antithetic variates is demonstrated in figure 4.8. The left and middle panels use simple random sampling but with different numbers of samples. Since the middle panel uses a thousand times more samples it should be the more accurate approximation. The left panel uses one thousand samples and has noticeable red regions, indicating negative approximations of $\mathbb{IKG}$. Even where the approximation is positive there are noticeable differences between the left and middle panels. On the other hand the right panel, which uses antithetic variates, closely matches the middle panel and has no negative regions. Even though the right panel also only uses one thousand samples, the antithetic variates technique allows it to achieve the same accuracy as the one million sample approximation.

Second, it was found that the majority of the expense in calculating the sampling approximation came from the calculation of the covariance $\tilde{\sigma}_n$. Because the value of $\tilde{\sigma}_n$ depends on t' but not on z, there is an advantage to using fewer unique t' in the samples. To take advantage of this, a smaller set of t' were generated and each was paired with multiple z draws to construct all the $\{t'_i, z_i\}$ pairs. A ratio of around 100 z samples for each t' sample was found to be effective.

Finally, as another variance reduction measure, stratified sampling was used for both t' and z. This ensures that representative samples come from each part of the distribution. To generalize the stratified t samples to higher-dimensional problems, Latin hypercube

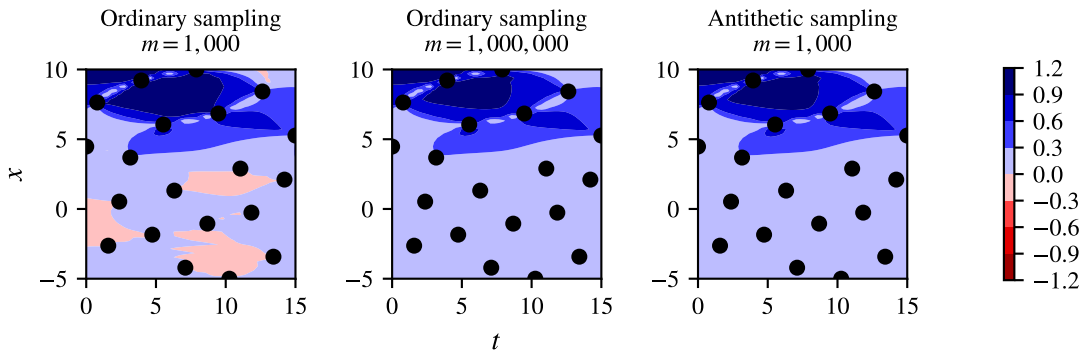

Figure 4.8: Effect of using antithetic variates to estimate $\mathbb{IKG}$. Each panel shows contours of approximate $\mathbb{IKG}$ for the Branin function using the same twenty maximin Latin hypercube data points. The left and middle panels use simple random sampling with one thousand and one million draws, respectively. The right panel uses antithetic variates of the standard normal variable z and one thousand samples.

sampling can be used. Note that this is ordinary Latin hypercube sampling and not a maximin Latin hypercube as would be used in a space-filling design of experiments.

4.4 Experiment 2: Artificial Test Functions

The Bayesian parameterized optimization method proposed in section 4.2 is novel, so experimentation is needed to validate that it is actually capable of solving the tow-steered composite technology performance estimation problem that inspired its creation and thereby support hypothesis 2.

Beiranvand, Hare, and Lucet [123] described best practices and provided a general framework for comparing optimization algorithms, which should largely translate to parameterized optimization and help to avoid known pitfalls when designing the experiments. A key insight is that domain-specific test problems, like the physics-based testbed from section 3.3, provide the most useful information about the method, but also have disadvantages. Domain-specific test problems are usually available in limited numbers, lack known solutions, and can be computationally expensive. In particular, Sala, Baldanzini, and Pierini [124] observed that while an algorithm's performance is most important on expensive

problems, it is also the most difficult to study. In contrast, representative artificial test functions perform better in these regards and provide a nice complement to domain-specific problems. This section describes an experiment that tests the performance of the Bayesian parameterized optimization method on a set of artificial test functions, and section 4.5 tests the method on the more cumbersome physics-based testbed.

4.4.1 Procedure

The steps in Beiranvand's framework for comparing optimization algorithms are: 1) clarify the reason for the comparison, 2) select the test set, 3) perform the experiments, and 4) analyze and report the results. For this work there is not a comparison per se, because no existing methods for parameterized optimization of expensive objective functions were found in literature. However, the goals of the experiment can still be identified:

- What are the characteristics of objective functions the method can succeed on?
- How much accuracy can be achieved with a given budget of objective function evaluations?
- Is it possible to know if the method is succeeding while it is running?
- If the method does not succeed, what compromises need to be made to the problem formulation so it can?

Test Functions

In order to determine objective functions that the method can succeed on, the test set needs a large number of functions with a variety of characteristics and "difficulties". There are many approaches for creating random test functions in literature, such as those presented by Ng and Li [125], Addis and Locatelli [126], and Gaviano et al. [127]. An advantage of these test functions is they have known optima when used as global optimization objective functions, but this desirable property breaks down if used as a parameterized optimization objective functions.

The test function generation for this experiment followed the approach used by Loeppky, Sacks, and Welch [93], who used functions drawn from Gaussian process priors with the squared exponential covariance from equation (4.7) as test functions while analyzing the accuracy of surrogate models. They found that the curse of dimensionality [128] hurt surrogate accuracy as expected, but also found strong effects caused by the sensitivity hyperparameter statistics τ and ψ, shown in equations (4.19a) and (4.19b). These can be used as tuning knobs to create functions with a variety of characteristics and difficulties.

$$\tau = \sum_{i=1}^{d} \theta_i \tag{4.19a}$$

$$\psi = \sum_{i=1}^{d} {\theta_i}^2 \tag{4.19b}$$

The sum τ can be thought of as the total sensitivity. As it increases the correlation between the responses at different locations decreases, which naturally makes the problem more difficult. The sum of squares ψ corresponds to the sparsity of effects, or how much the function's variation is concentrated in particular dimensions. Figure 4.9 shows a comparison of functions drawn from Gaussian processes with the same τ but different ψ. As ψ increases with constant τ the function increasingly varies in the horizontal direction and stops varying in the vertical direction. Higher effect sparsity makes it easier to make an accurate surrogate model, and in the most extreme case all of the function's variation is only associated with a single dimension.

Equation (4.20) can be used to convert a desired number of dimension d, τ and ψ into a vector of sensitivity hyperparameters. This equation is in terms of b rather than ψ so a root finding algorithm is needed to adjust b to the proper value.

$$\theta_i = \tau \left[\left(1 - \frac{i-1}{d}\right)^b - \left(1 - \frac{i}{d}\right)^b \right] \quad \text{for} \quad i = 1, ..., d \tag{4.20}$$

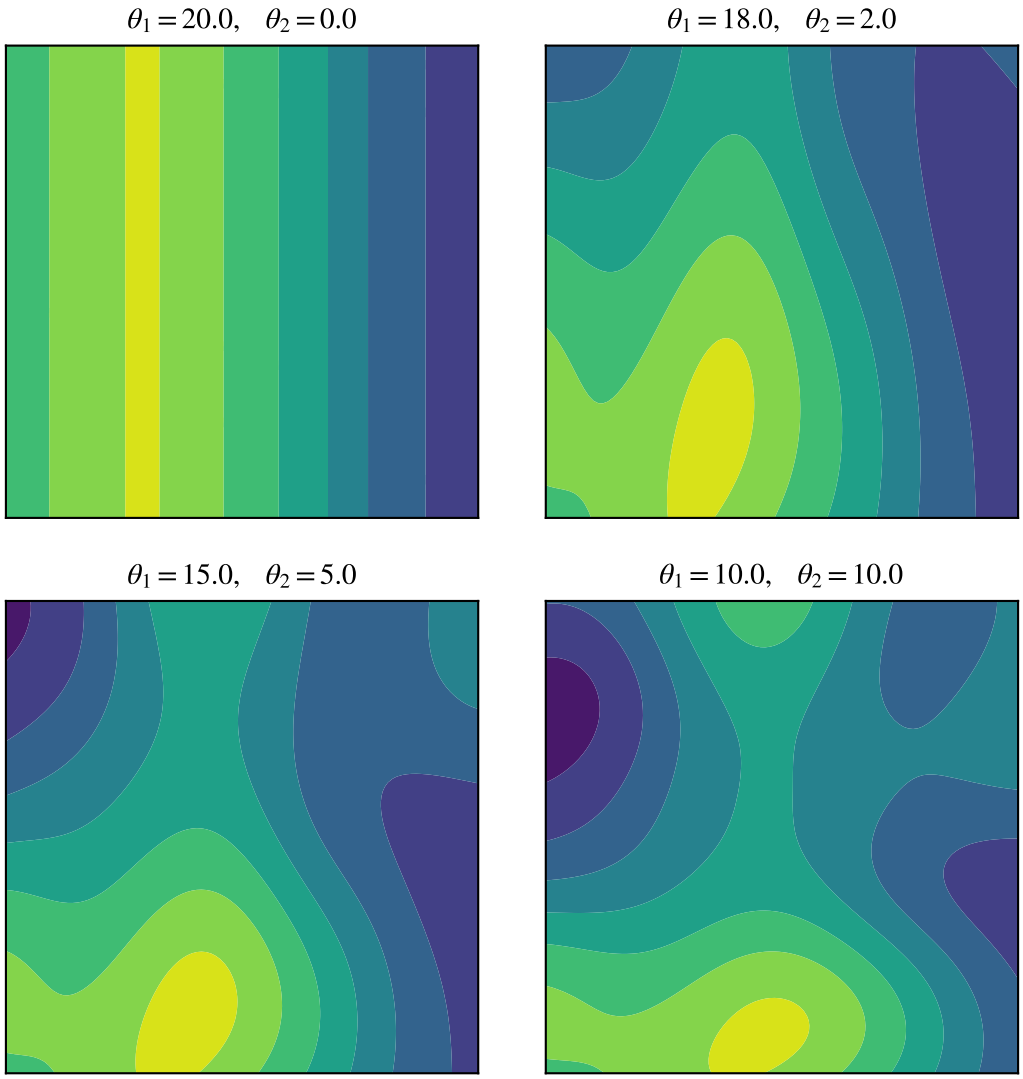

Figure 4.9: Test function sparsity of effects. Though all functions have the same total sensitivity, the allocation of sensitivity between the two dimensions differs. The functions vary more rapidly in the horizontal direction than in the vertical direction, except for the bottom right function which has an even distribution of sensitivity.

The test functions need to resemble the behavior of the intended application in order to provide relevant information. Gaussian processes with the squared exponential covariance function are a standard choice in engineering applications, and embody the reasonable assumption that the response smoothly varies as a function of the inputs, which somewhat justifies their use here. Using a range of hyperparameters makes it more likely that at least one of the hyperparameter settings will provide similar behavior to the domain-specific problem. However, it is important to note that the actual physics-based analysis might produce a response that is somewhat rougher or noisier-appearing than these test functions. It is also important to be aware of the fact that the assumed model form is also used to generate the test functions so there is no possibility of model misspecification, which is a legitimate concern for real data sets. These simplifications were made deliberately in order to focus this experiment on the behavior of the Bayesian parameterized optimization method itself, by studying it under somewhat ideal conditions. The next section will ensure that these practical considerations are addressed and cover the topic of Gaussian process covariance selection.

Factors and Experimental Design

The factors and ranges studied in the experiment are shown in table 4.1. All but one of the factors control the characteristics of each case's objective function and the remaining factor controls the total number of samples taken to construct a surrogate model.

Since a primary goal of the experiment is to determine the effect these factors have on the Bayesian parameterized optimization method's performance, and because noisy results are likely due to the randomness of the artificial objective functions, a classical design of experiments was used. A central composite design with 80 cases was selected so that all main, pure quadratic, and two-factor interaction effects could be resolved.

The factors controlling the objective function consist of the number of dimensions and the hyperparameter statistics identified above, but because the parameterized optimization

problem has two distinct sets of variables these factors are further refined so that the parameter and choice variable subsets can be varied independently.

The total number of dimensions ranged from four to twelve in order to cover dimensionalities where surrogate-base approaches are expected to work. The dimension settings are all divisible by four so that the parameter dimension fraction can neatly divide them into parameter dimensions and choice variable dimensions with fourth-based ratios.

The total sensitivity covers a range that includes both easy and difficult objective functions [93], and the amount of sensitivity allocate to each variable subset is controlled by the parameter sensitivity fraction.

The effect sparsity for each variable subset is defined on a normalized $[0, 1]$ range to decouple it from the total sensitivity (the possible values of ψ are constrained by the value of τ). At zero, there is minimum effect sparsity and all individual sensitivities are the same. At one there is maximum effect sparsity and only one dimension's sensitivity is nonzero. Since this would effectively create a one-dimensional objective function, the range for these factors did not go all the way to one.

Finally, the initial sampling in each case uses the hueristic of $10d$, and the sequential sampling technique is used to bring the total number of points up to between $11d$ and $15d$.

An idiosyncrasy of these factors is that the effect sparsities are not unique when there is only one dimension in their respective variable subset. This happens when there are four dimensions and the dimension fraction is either $1/4$ or $3/4$. In that case, all of the variable subset's sensitivity goes into the single dimension regardless of the case's effect sparsity setting. This redundancy is not ideal but it allows for the experimental range to be a neat hypercube and avoids the need for a custom design of experiments that honors linear constraints.

Because of this redundancy, there should be an interaction between the number of dimensions, the parameter dimension fraction, and the effect sparsities such that the effect sparsities have no impact when they are only associated with one dimension.

Table 4.1: Factors and ranges for artificial test function benchmark experiment

Factor	Description	Settings		
		Low	**Medium**	**High**
d	number of dimensions	4	8	12
d_t/d	parameter dimension fraction	$1/4$	$1/2$	$3/4$
τ	total sensitivity	1	5.5	10
τ_t/τ	parameter sensitivity fraction	$1/4$	$1/2$	$3/4$
$\frac{d_t}{d_t-1}\left[\frac{\psi_t}{\tau_t^2}-\frac{1}{d_t}\right]$	parameter effect sparsity	0	$1/3$	$2/3$
$\frac{d_x}{d_x-1}\left[\frac{\psi_x}{\tau_x^2}-\frac{1}{d_x}\right]$	choice variable effect sparsity	0	$1/3$	$2/3$
n/d	samples per dimension	11	13	15

Process

Each case started by translating the factor settings into covariance rate hyperparameters using equation (4.20), and then into a random function drawn from the specified Gaussian process prior. The functions were implemented as the weighted sum of 10,000 sinusoidal features, using the process described in appendix B of Hernández-Lobato, Hoffman, and Ghahramani [119]. Their domain was the unit hypercube $[0, 1]^d$. After the function was created, two different surrogate models were created: the first collected data using the Bayesian parameterized optimization method and the other, serving as a baseline, used a maximin Latin hypercube. This baseline approach was chosen to be representative of an attempt to create a globally accurate surrogate model since a competing method for parameterized optimization of expensive objective functions was not found.

The process for creating the Bayesian parameterized optimization surrogate started by evaluating a maximin Latin hypercube with $10d$ points on the true function. Then data points were iteratively added using algorithm 2 until the specified amount had been obtained. No hyperparameter estimation was performed, each iteration's surrogate was given the true hyperparameters (hyperparameter estimation is part of the experiment described in the next

section). Using the current surrogate, the next point to add was found by maximizing the approximate $\mathbb{IKG}$ acquisition function with 1,000 Monte Carlo samples and $10d_t$ unique parameter vectors. Antithetic variates and Latin hypercube sampling were used to improve the approximation, as described in section 4.3. The maximization used multistart gradient-based optimization with $10d$ starts.

The control surrogate was simply constructed by evaluating a maximin Latin hypercube with the specified number of points. Like the adaptive surrogate, this one was also given the correct hyperparameters.

Responses

The responses recorded in each case are measures of the error in the surrogate model value functions. These were empirically estimated by sampling $10d_t$ vectors from the parameter space, finding the optimal choice variables at each location, and comparing the final surrogate's optimal value to the true function's optimal value. For both surrogates, the error samples were normalized by the range of the true value function and aggregated into a maximum error, root mean square error (RMSE), and bias.

Ideally the optima of the true function would have been known, but a technique for achieving this for a parameterized optimization problem was not found or devised. However, since the true function is inexpensive to evaluate, it is not too troublesome to thoroughly search for the true optimum using a global optimization technique. In this case the basin-hopping algorithm [129] implemented in SciPy [84] was used.

4.4.2 Results and Discussion

A scatterplot matrix of the responses is shown in figure 4.10, with the single-stage error on the left side and the adaptive/multistage error on the right. All of the responses are best shown on a log scale, though the sign of the bias can be negative. The two sampling strategies exhibit similar trends: RMSE and maximum error are highly correlated, and both

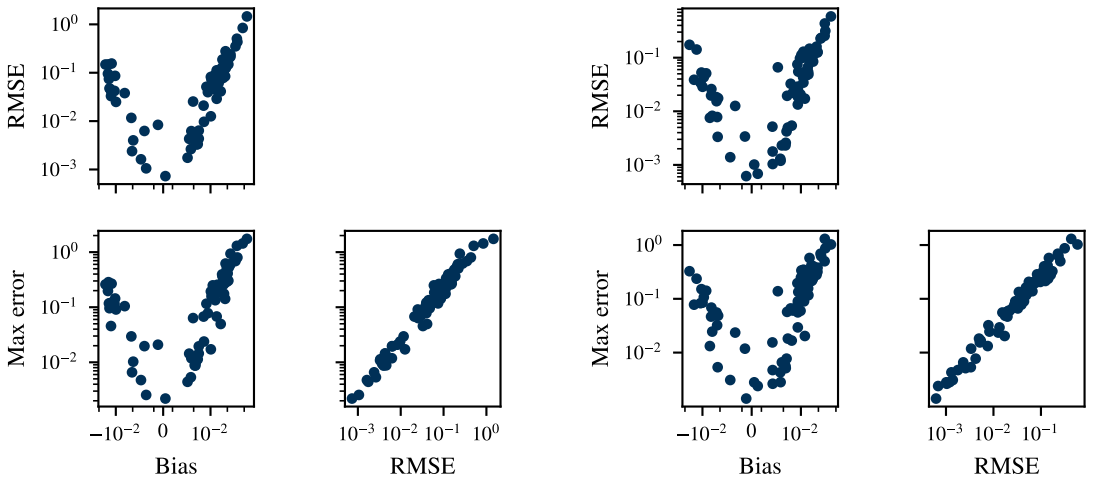

Figure 4.10: Scatterplot matrix of artificial test function experiment responses. The left group shows the responses for the surrogates constructed from a single-stage Latin hypercube and the right group shows the responses for the surrogates constructed using multistage Bayesian parameterized optimization.

RMSE and maximum error are highly correlated with the absolute value of the bias. Since the three responses are so strongly related, subsequent analysis focuses on just the RMSE and all findings are assumed to also apply to maximum error and bias.

Before setting the bias aside, figure 4.11 shows that it tends to be positive, especially when its magnitude is large. Therefore, it can be expected that the surrogate value function will tend overestimate the true value function.

Figure 4.12 shows a comparison of the raw RMSE for both kinds of sampling strategies. The two responses are strongly correlated, indicating that the characteristics of the objective function have a strong effect on surrogate model accuracy regardless of how the samples are collected. Nonetheless, a strong majority of the points lie to the bottom right of the line, which indicates that the Bayesian parameterized optimization method is performing better than the baseline space-filling design of experiments. Because the data is on a log scale, the amount of improvement is larger than it might appear at first glance.

In order to provide a more quantitative analysis, a couple of quadratic response surface

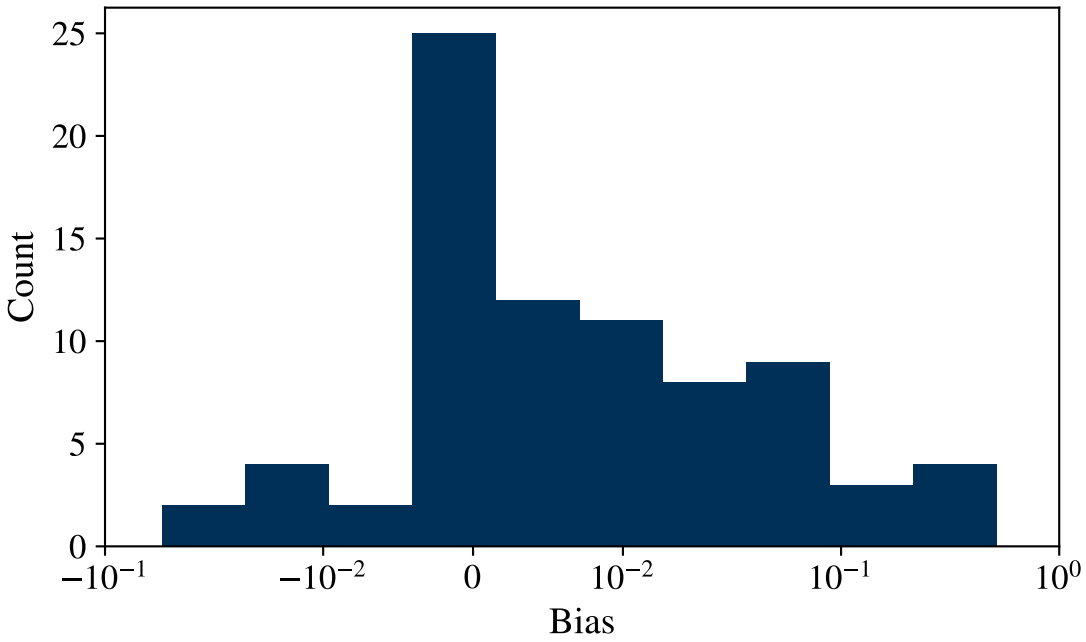

Figure 4.11: Histogram of error bias in artificial test function experiment for Bayesian parameterized optimization surrogate

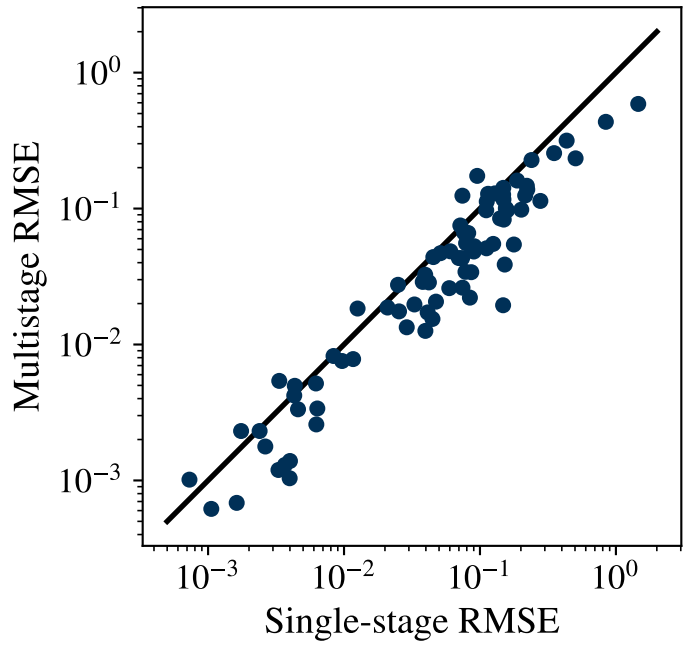

Figure 4.12: Comparison of single-stage and multistage surrogate RMSE

models were fit to the data. The first model fit the RMSE of the multistage surrogate model in order to determine the effects that the objective function characteristics and number of samples has on the accuracy, and the second fit the ratio of multistage RMSE to single-stage RMSE in order to measure the improvement relative to the benchmark.

RMSE Analysis

Because of the log scaling of the data the RMSE was log transformed before being fit by the model. Initially, the data was fit using standard least squares regression and achieved a high R^2 of 0.95. However, the PRESS R^2, a cross-validation measure, dropped to 0.84 indicating that this model might be overfitting.

Another model was fit using adaptive lasso regression. Lasso regression guards against overfitting by using an ℓ_1 penalty term to regularize the maximum likelihood model parameter estimates. This both reduces variance of the estimates and forces some of them to zero, essentially performing model selection. An appropriate size for the penalty term was determined by using AICc as a validation criteria. The R^2 for this more trustworthy model was still quite high at 0.94. The predictive ability of this model, though on a log scale, seems surprising when considering the amount of randomness involved in each experimental case. In each case: 1) an artificial test function was randomly generated, 2) the first-stage maximin Latin hypercube was randomly generated (and the maximin Latin hypercube for the baseline approach), and 3) the approximate $\mathbb{IKG}$ in each iteration is evaluated using Monte Carlo sampling.

The typical RMSE across the experimental space was 0.083, about 10% error. The strongest effects found by the model are listed in table 4.2. The middle column shows the size of the scaled parameter estimate and the right column shows the percent change in RMSE that the effect typically causes across the design space. The strongest effects were the main and quadratic effects from the total sensitivity, which can change the RMSE by more than an order of magnitude across the experimental range. The model also reported

Table 4.2: Strongest effects on Bayesian parameterized optimization surrogate RMSE

Factor	Effect size	Impact
total sensitivity	0.548	+1150%
dimensions	0.205	+157%
cases per dimension	-0.205	-61%
choice variable effect sparsity	-0.195	-60%
parameter effect sparsity	-0.158	-52%
parameter dimesion fraction	-0.087	-33%
total sensitivity $\times$ total sensitivity	-0.451	-87%
parameter dimension fraction $\times$ parameter effect sparsity	-0.146	-49%
parameter dimension fraction $\times$ choice variable effect sparsity	0.114	-41%
total sensitivity $\times$ parameter effect sparsity	0.064	+34%

strong effects from the number of dimensions and the number of data points collected per dimension. As expected more dimensions hurt surrogate accuracy and more data helps.

There were also similar sized effects for the effect sparsities indicating that more effect sparsity helps. The effect sparsities also have similar sized interaction effects with the parameter dimension fraction. These interactions have opposite signs because the parameter dimension fraction has opposing effects on the size of the parameter and choice variable subsets. The intuitive explanation of these effects is that the effect sparsity is more important in whichever variable subset is larger. Finally, there are smaller effects associated with the parameter dimension fraction. It suggests that parameterized optimization problems with fewer parameters and more choice variables are more difficult to solve using this approach.

This regression model provides clear information on how the characteristics of the objective function and the number of cases affect the accuracy of the resulting surrogate model.

RMSE Ratio Analysis

The second response surface model was fit to the ratio of the multistage surrogate RMSE and the single-stage surrogate RMSE, again on a log scale. The purpose of this model is to show the improvement in surrogate model accuracy that Bayesian parameterized optimization provides compared to the baseline approach. Like the previous model, this one was fit using adaptive lasso with AICc validation. The original linear least squares fit for this model had a negative PRESS R^2!

The fit for this model was noticeably worse with an R^2 of 0.61, though there are still meaningful trends. The typical advantage of Bayesian parameterized optimization compared to the baseline was a 35% reduction in RMSE. The largest parameter estimates from the model are shown in table 4.3. One of the strongest beneficial effects was from the increasing the number of cases, which makes sense because at the lowest setting only about a tenth of the data is chosen adaptively while at the highest setting about half of the data is. This effect is very important because it suggests that the improvement will continue to grow as the number of cases increases beyond the $15d$ bound used in this investigation. There is also a small interaction effect between number of cases and the parameter dimension fraction showing that the proposed method is more suited to problems with few parameters and many choice variables, like the tow-steered composite problem.

There are almost equal effects from the parameter dimension fraction and the parameter sensitivity fraction that show that there is more advantage from Bayesian parameterized optimization when the objective function is more difficult in the choice variables. Additionally, the effect from choice variable sparsity shows that the adaptive method is better able to exploit objective functions when the choice variables exhibit sparsity of effects.

Even though this model fits worse than the previous and has a lower R^2, that just means that the factors have less of a role in determining the response. The key takeaway from this model is that in the typical case Bayesian parameterized optimization provides a 35% reduction in error compared to the baseline. The poor fit only makes it a bit harder to

Table 4.3: Strongest effects on Bayesian parameterized optimization advantage

Factor	Effect size	Impact
cases per dimension	-0.062	-25%
parameter dimension fraction	0.047	+24%
parameter sensitivity fraction	0.046	+24%
choice variable sparsity	-0.037	-15%
dimensions $\times$ choice variable effect sparsity	0.073	+40%
parameter sensitivity fraction $\times$ parameter effect sparsity	-0.054	-22%
parameter dimension fraction $\times$ choice variable sparsity	0.044	+22%
parameter dimension fraction $\times$ cases per dimension	0.032	+16%

determine how that will change based on the characteristics of the objective function.

Conclusions

The results of this experiment give reason to believe that Bayesian parameterized optimization could be an effective way to gather data for building a surrogate model of tow-steered composite structural performance, or for that matter other parameterized optimization problems found in engineering that have an expensive objective function. The results show that the method is able to prioritize more favorable regions of the design space in a way that allows the solution set to be found with less evaluations than would otherwise be needed. If this behavior hold, it would provide the desired improvement in the trade-off between technology benefit and computational cost that has been identified as important to the thesis methodology.

The models of the results can be used to infer how much surrogate model error is likely and the relative performance of Bayesian parameterized optimization. However, the results also show that if the function is too difficult, the use of Bayesian parameterization will *not* be the deciding factor that turns an intractable problem into a tractable one; some compromise,

like shrinking the design space, would be needed. In order to provide more confidence that this technique can be successfully applied to the tow-steered composite design problem, it still needs to be tested on a domain-specific problem.

4.5 Experiment 3: Physics-Based Testbed

The results of the previous experiment show that the proposed Bayesian parameterized optimization method is an effective way to collect data to make a surrogate model for objective functions with a variety of characteristics. However, because it used test functions sampled from the assumed model form, and because the characteristics of those test functions that most closely resemble physics-based analysis are unknown, there is still uncertainty whether the method would actually be effective on the tow-steered composite technology performance estimation problem. The experiment in this section applies the Bayesian parameterized optimization method to a more realistic objective function formed from the physics-based testbed from section 3.3. In doing so it confronts the practical difficulties of using such a method, like model selection and hyperparameter estimation, that were disregarded in the previous experiment. This will allow for a more conclusive test of hypothesis 2.

Problem Formulation

The physics-based testbed described in section 3.3 was cast as a parameterized optimization problem's objective function. The parameters were the planform area and the aspect ratio, and the choice variables were 7 B-spline coefficients controlling the tow orientations on each of the upper and lower skins. This was a total of 16 input variables for the surrogate model, divided into 2 parameters and 14 choice variables. The ranges for theses variables are shown in table 4.4. The wing area is extended 50% above and below the baseline value, and the aspect ratio is extended 50% above the baseline value. The tow path orientations were allowed to vary between -60° and 60°.

Table 4.4: Variable ranges for surrogate modeling of physics-based testbed

Group	Variable	Description	Settings	
			Low	**High**
Parameters	S	planform area	450,000 in^2	750,000 in^2
	AR	aspect ratio	9	13.5
Choice variables	$\theta_{upper,i}$ $i \in [1,7]$	upper coefficients	-60 deg	60 deg
	$\theta_{lower,i}$ $i \in [1,7]$	lower coefficients	-60 deg	60 deg

As per the conclusions from chapter 3, the structural panel thicknesses were set by solving a lower-level optimization problem. Like the tow paths, the thicknesses for each of the five structure groups was defined by a B-spline with 7 coefficients. For this particular application, because gradients were available for this simple physics-based testbed, the thicknesses were set by a gradient-based optimization algorithm rather than fully stressed design.

Initial Data Collection

The initial stage of data was collected using a 160-case maximin Latin hypercube. An additional Latin hypercube with 40 cases was evaluated to serve as a holdout data set for validating the surrogate models of the objective function. Finally, the holdout set was directly optimized with a gradient-based algorithm in order to generate a third data set that provided a comparison point for the surrogate value function. These directly optimized cases were started with all tow orientations pointing straight along the span of the wingbox.

4.5.1 Model Selection

Alternatives

After collecting the initial data, the next step in the method is to select and validate the surrogate model. Multiple alternative model forms were evaluated since it was not known

Table 4.5: Modeling alternatives

Attribute	Alternatives		
mean function form	constant	linear	quadratic
covariance function form	squared exponential	Matérn 5/2	Matérn 3/2
factor analysis directions	0	1	2
hyperparameter estimation	MLE	REML	

which would best represent the tow-steered composite structural weight data. The options are enumerated in table 4.5. The choices for each attribute are independent of each other so there were a total of 54 different combinations. An advantage of an expensive objective function is that it was comparatively inexpensive to try all these different model forms.

The two components of a Gaussian process are its mean function and its covariance function. The alternatives considered for the mean function were a constant mean, a linear mean, and a quadratic mean. The mean function plays a similar role to a linear least squares model, with the Gaussian process accounting for additional nonlinear behavior. The quadratic alternative only used pure quadratic terms, as a full quadratic model with all interactions between the 16 inputs would have had total of 153 terms, and there was not enough data to reliably fit all of them. As it was, the quadratic alternative had 33 terms (1 intercept, 16 linear terms, and 16 pure quadratic terms).

There were two attributes that affected the covariance function. The first was the form of the covariance function. The options were the classic squared exponential, the Matérn 5/2, and the Matérn 3/2. The Matérn covariance family generalizes the squared exponential, and the resulting Gaussian processes become less smooth moving forward through this list. The less-smooth alternatives were considered because there are several reasons the data from the physics-based testbed might appear rough or noisy; these reasons include tolerances on numerical calculations, discrete changes in the structural topology, and discrete changes in the mesh.

The other choice affecting the covariance was the number of factor analysis directions. The covariance function essentially assigns a relatedness to input locations based on their distance from each other. Factor analysis directions give the covariance the flexibility to rotate the principle directions of the distance measure off of the coordinate axes. This flexibility comes with a possibility of overfitting the data, so only two factor analysis directions at most were allowed. If these two were found to help with model performance, then more could be added and evaluated.

The final model choice was how the hyperparameters were estimated. The hyperparameters consist of all the coefficients used in the mean function, a covariance magnitude, a covariance rate or length scale for each input dimension, any factor analysis directions, and a noise magnitude. The first option is the commonly used maximum likelihood estimation (MLE), and the other is restricted maximum likelihood (REML). REML estimates the hyperparameters in two steps by first estimating the covariance hyperparameters in a projected space orthogonal to the mean function's coefficients, and then estimating the mean function's coefficients using those covariance hyperparameters. The idea is that by estimating the covariance hyperparameters independently of the mean, the bias of the estimation and the chance of overfitting is reduced.

Results

All of the model alternatives were fit to the data using the specified hyperparameter estimation technique. Because of the well-known possibility of multiple local likelihood maxima, multistart gradient-based optimization was used. After estimating the hyperparameters the models were evaluated based on the training log likelihood, validation log likelihood and RMSE, and leave-one-out cross-validation log likelihood and RMSE. The RMSE only measures how close the model's mean prediction comes to the actual value, while the log likelihood accounts for both how close the mean prediction is as well as how confident the model is. The leave-one-out cross-validation used the hyperparameter estimates that were

obtained from all the data, but then held out each individual case to make predictions. Note that these metrics measure how well the surrogate model represents the objective function and not the value function, since in a real application the value function would be unknown and impractical to sample from.

The most immediately apparent result was how badly the models that used factor analysis directions could be. The validation log likelihood for the 18 models with no factor analysis directions ranged between -15.3 and -3.2, and the worst three models with factor analysis directions had validation log likelihoods of -8,300, -8,600, and -200,000! These models severely overfit the training data and could only be exposed as overconfident by examining their predictions for the validation set. Using REML instead of MLE mitigated the overfitting but the three worst models trained with REML still had poor validation log likelihoods of -220, -480, and -1,500.

Figure 4.13 compares the validation metrics with the leave-one-out cross-validation metrics. Since the validation metrics were calculated on an independent holdout set, while the leave-one-out cross-validation metrics were calculated on the same data used to estimate the hyperparameters, the validation metrics should be more representative of each model's predictive ability. Because of the extreme skew in the validation log likelihoods only models with no factor analysis directions were included in the log likelihood plot. These plots show that the leave-one-out metrics are negatively correlated with the validation metrics. This is surprising and shows the importance of having an independent set of validation data when selecting the model form, as in this circumstance leave-one-out cross-validation provided actively misleading results.

Figure 4.14 compares the validation RMSE and log likelihood. There is not a significant trade-off between these criteria as better RMSE and better log likelihood are strongly correlated. It was decided to use validation log likelihood as the ultimate criterion for selecting the model form since it also accounts for whether the model is appropriately confident in its prediction.

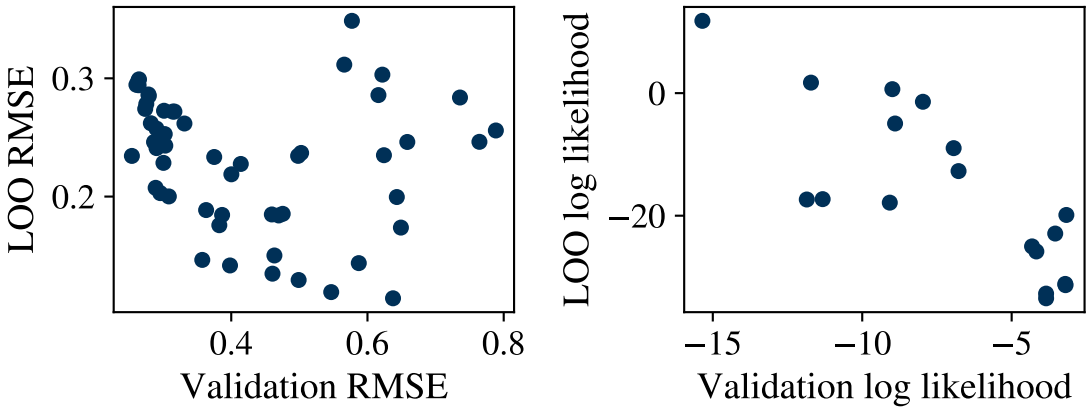

Figure 4.13: Performance of model alternatives. The plots compare the validation performance to the leave-one-out cross-validation performance, measured by root mean square error in the left pane and log likelihood on the right. Due to the extreme skew, the log likelihood only shows results for models with no factor analysis directions.

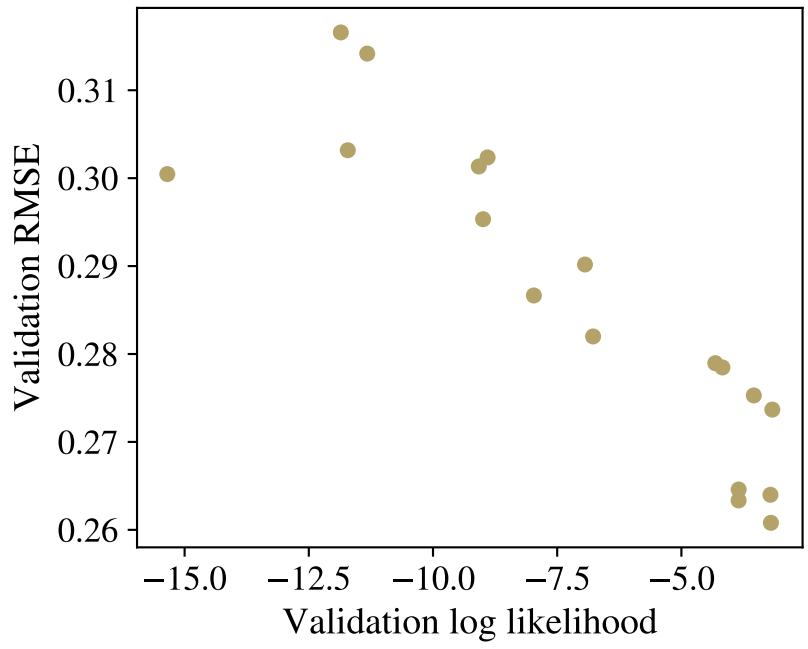

Figure 4.14: Performance of model alternatives on validation data. Due to the extreme skew only results for models with no factor analysis directions are shown.

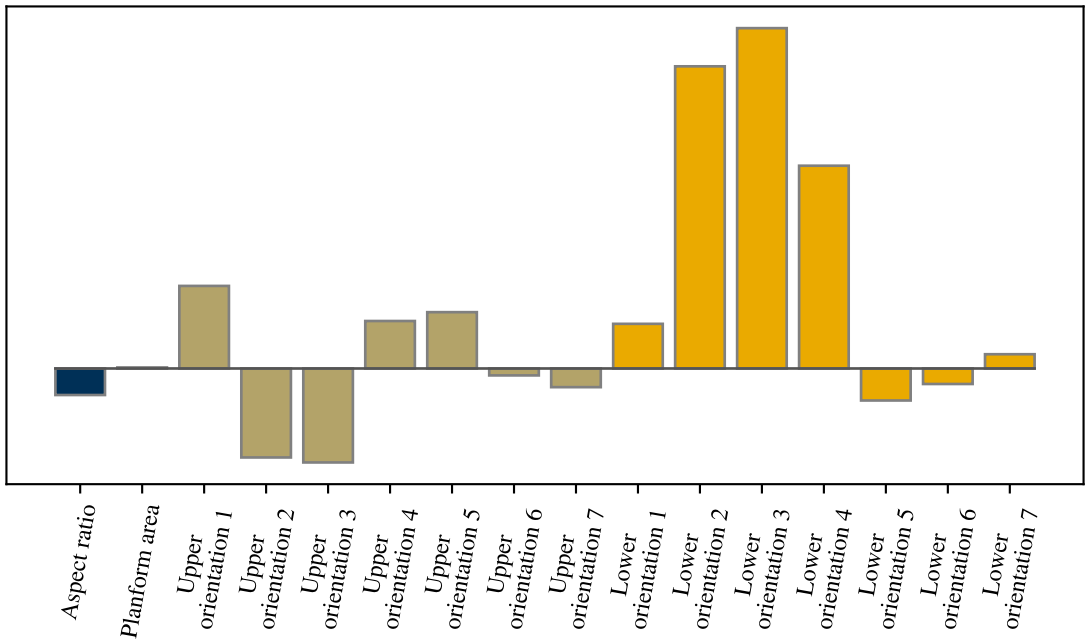

Figure 4.15: Largest eigenvector of best-performing model's distance metric

The model with the best validation log likelihood actually had a factor analysis direction. This represents a direction that is particularly important for explaining the similarity between the responses at different locations. Figure 4.15 shows the largest eigenvector of this model's distance metric, which is primarily composed of the tow orientation angles of the middle lower skin. This means that, for the variable ranges considered, this model interpreted wingbox designs with different tow orientations in this region as being the least related to each other.

The best-performing model also had a linear mean function, a Matérn ⅝ covariance function, and used MLE for hyperparameter estimation. Because some of the other models employing factor analysis directions performed so terribly, and because some models that used no factor analysis directions came close to matching the best validation log likelihood, the decision was made to not use factor analysis directions. Even though they provided a benefit in this particular instance, in general they were associated with a high risk of

overfitting.

In order to understand the trends associated with the modeling alternatives, the validation log likelihood was fit to a main effects and two-factor interaction model of the mean, covariance form, and hyperparameter estimation technique. The results indicated a benefit to using a constant or linear mean rather than a quadratic mean, and a slight advantage to using one of the Matérn covariances rather than the squared exponential. Based on these trends, the decision was made to use a constant mean, a Matérn 5/2 covariance with no factor analysis directions, and to use MLE to estimate the hyperparameters. The Matérn 5/2 covariance was chosen over the Matérn 3/2 because it is smoother, enabling the surrogate model to make stronger inferences from the data. When this model form was evaluated it achieved a validation log likelihood of -3.2 which is similar to the best model's validation log likelihood of -0.3.

The initial stage actual vs. predicted plot for the chosen model is shown in figure 4.16. The predictions and uncertainty for the training set were calculated using a leave-one-out procedure. This plot shows that the initial surrogate model was able to make predictions with limited accuracy, motivating the use of Bayesian parameterized optimization to improve the surrogate. The model did not seem to predict the training data better than the validation data, so it was not overfit. Figure 4.17 shows the model's normal quantiles plot: the scaled prediction residuals were approximately normal and the model had a good understanding of its own accuracy. Since this initial model was demonstrated to be working correctly the data collection proceeded to the adaptive portion.

4.5.2 Sequential Data Collection

Following algorithm 2, new training data were added until the total number of points was doubled. The approximate calculation of $\mathbb{IKG}$ used 1,000 random samples with antithetic normal variates and a Latin hypercube of 20 unique parameter values.

It proved impossible to have any degree of confidence in finding the global maximum of

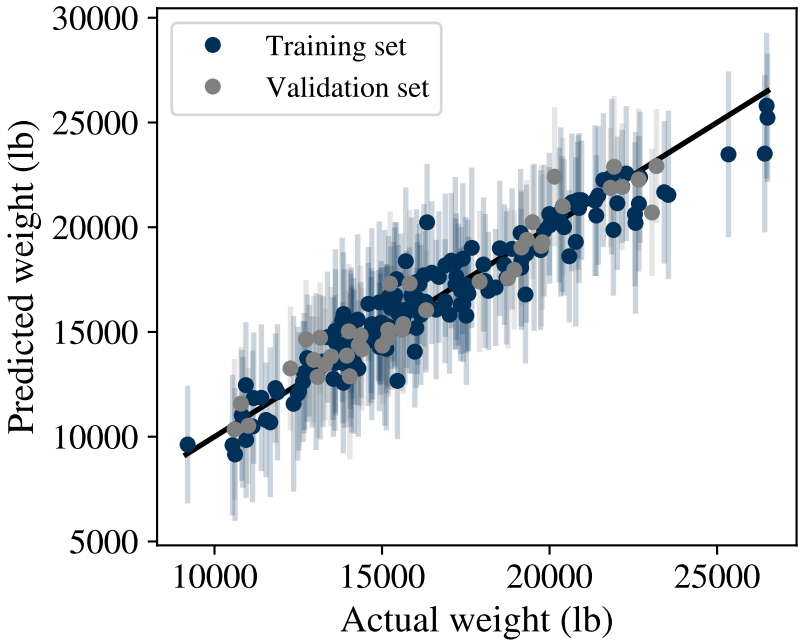

Figure 4.16: Initial stage actual vs. predicted plot of selected model. The error bars represent three standard deviations above and below the prediction mean.

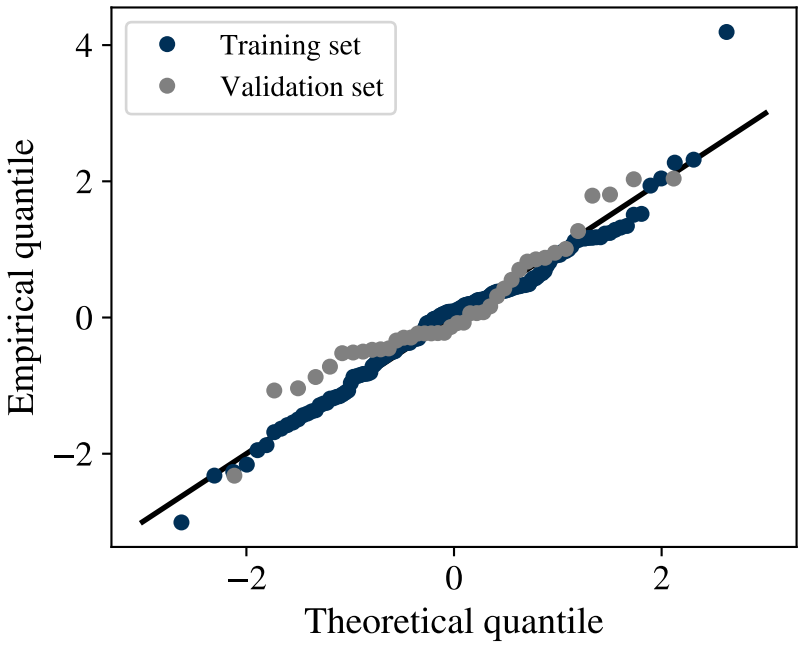

Figure 4.17: Initial stage normal quantiles plot of selected model

$\mathbb{IKG}$ so a heuristic approach was used. To choose the next point, 320 random locations were sampled and their $\mathbb{IKG}$ was evaluated. The 32 best-performing points from this preliminary screening were refined using a gradient-based optimization algorithm, and then the best one was selected. Typically all of the refined solutions settled in different local maxima, pointing to a vast number of local maxima and very complicated behavior of $\mathbb{IKG}$ in this number of dimensions. However, the several top-performing solutions in each iteration had similar values so it seems likely that the losses from not finding the true maximum are small. After selecting the new point to add to the training data, it was evaluated on the physics-based testbed and then the surrogate model was updated.

4.5.3 Results and Discussion

The final actual vs. predicted plot after collecting all the additional data is in figure 4.18, and shows several signs that Bayesian parameterized optimization worked as intended. First, the sequentially added data tend to have lower weights than the initial data which demonstrates that the new data is sampled from better-performing regions of the design space. Second, there is still a significant range of weights in the new data which is important because the algorithm should be exploring the parameter space and the optima from different parts of the parameter space have different weights. Third, the error bars associated with the new data are much smaller than those for the initial data; on the scale of the plot many are hardly even visible. This indicates that the new points are concentrated and that the surrogate now has much less uncertainty in the promising regions of the design space.

The method's exploration of the parameter space is further illustrated by figure 4.19, which displays the locations of the points added in the adaptive stage as well as a kernel density estimate giving the broad sampling behavior. Although there was a higher concentration of points at lower aspect ratio and planform area, samples covered the entire range of parameter space. The shape of the outermost contour is somewhat rectangular, resembling the shape of the design space. Also of note is that the contours tend to be wider in aspect

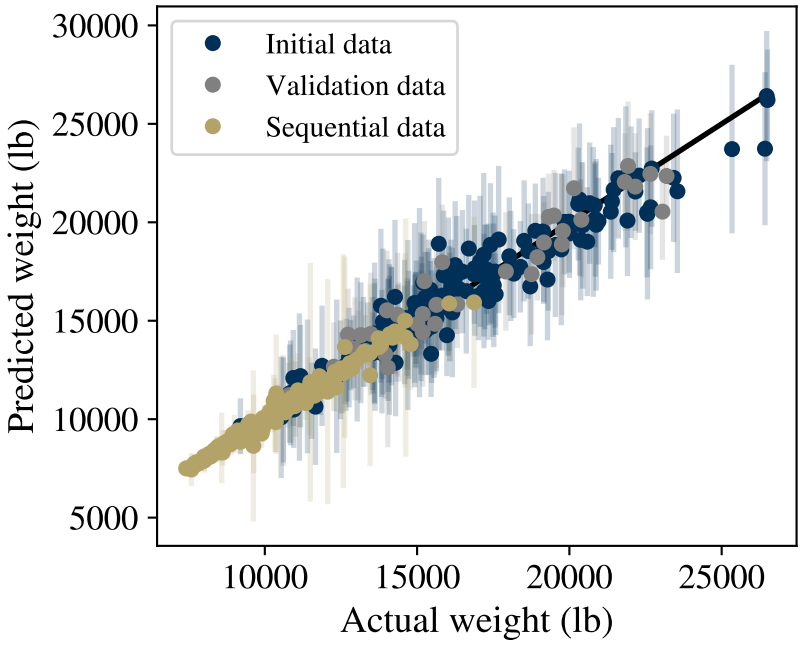

Figure 4.18: Final actual vs. predicted plot. The upper and lower error bars cover three standard deviations of the prediction distributions.

ratio than in planform area.

In contrast, figure 4.20 shows the selected samples in two of the choice variable dimensions corresponding to the upper skin tow orientation near the root and near the tip of the wing. Although the samples also cover the entire range of each variables, they are much more concentrated than in the previous figure. This image exemplifies how Bayesian parameterized optimization is able to improve surrogate model accuracy by focusing on the best-performing parts of the choice variable space. The dense region in the figure should cover the tow orientation settings that provide the best performance across all the possible planforms.

There is more insight into this sampling behavior, and the physics-based testbed itself, in figure 4.21. This bar graph shows the rate hyperparameters of the final surrogate model. Each rate can be thought of as the nonlinearity of the associated dimension across its range, and the more nonlinearity there is the more difficult it is to make predictions. The most

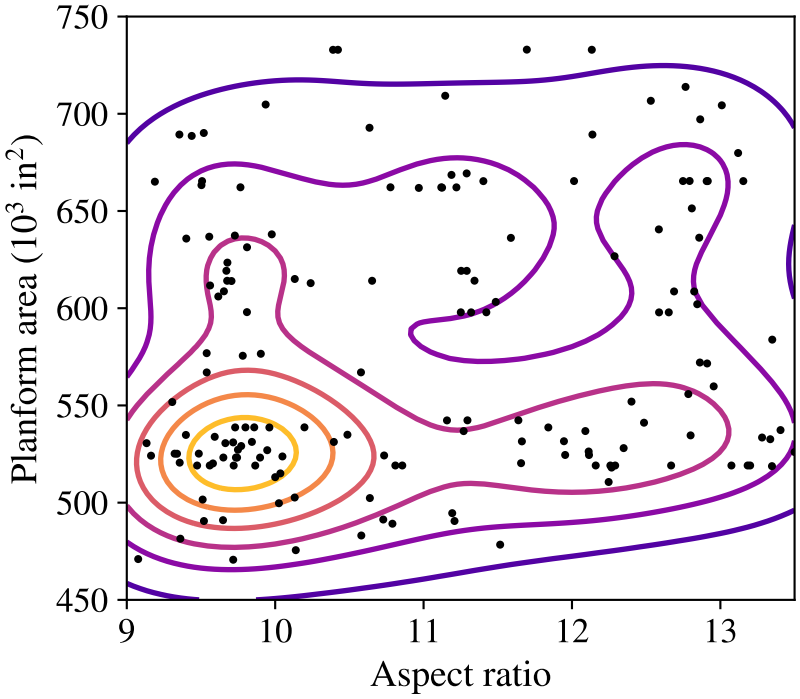

Figure 4.19: Adaptive samples in parameter space with kernel density estimate

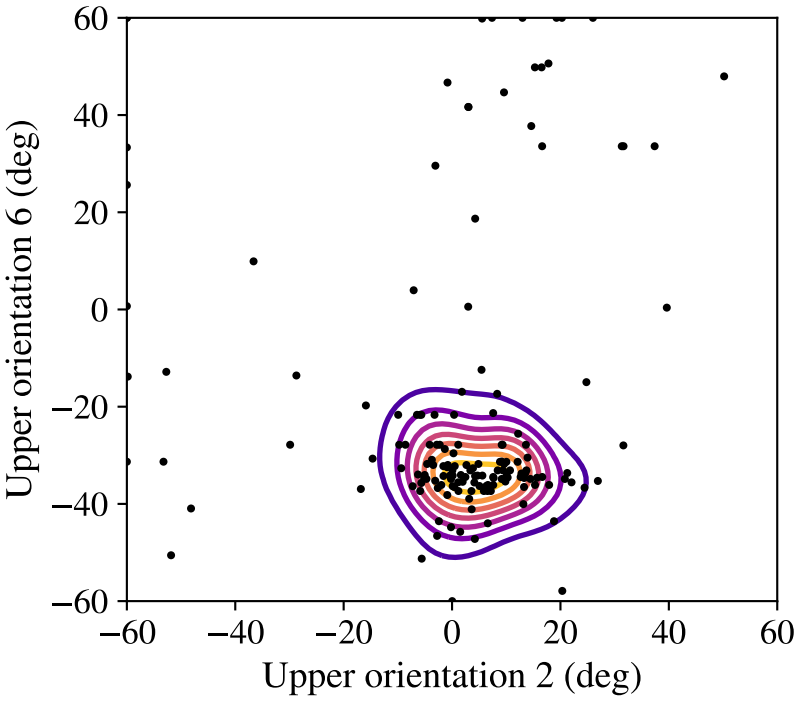

Figure 4.20: Adaptive samples in choice variable space with kernel density estimate

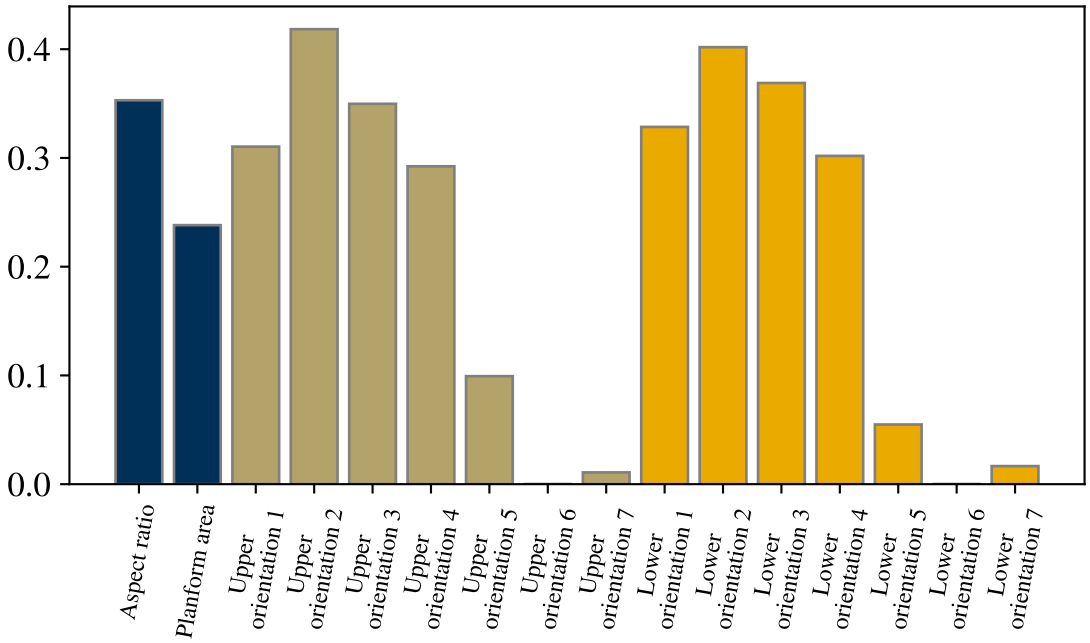

Figure 4.21: Scaled rate hyperparameters of final surrogate model

nonlinear dimensions are associated with the root and midspan tow orientations, and the least is associated with the wingtip tow orientations. There is also significant nonlinearity associated with the planform parameters, more so for aspect ratio than for area.

Looking back at figure 4.19, there was a sensible reason for the samples to be more uniform over the aspect ratio and more centralized over the planform area. Because the rate for aspect ratio is larger, it is more difficult to make predictions from distant samples. As a result, that dimension needs to be sampled more uniformly. For planform area the rate is smaller so it is easier to make predictions from distant samples; therefore the samples should be more centralized because interior samples provide information that applies to more of the total design space. The two dimensions in this particular example do not have drastically different rates, which is why this effect is noticeable but not drastic. This same effect can also be seen in figure 4.20, where the higher-rate upper orientation 2 has a wider distribution than the lower rate upper orientation 6.

Effectiveness of Bayesian Parameterized Optimization

In order to track the progress and quantify the effectiveness of the Bayesian parameterized optimization method, in each iteration the surrogate model's value function was compared to the control group of directly optimized wingbox designs that their thicknesses and tow paths set by a gradient-based optimization algorithm. After evaluating the latest point and updating the surrogate model, the surrogate was optimized at the same locations in parameter space as this comparison data.

Figure 4.22 shows the evolution of this surrogate value function error along with the integrated knowledge gradient for the chosen sample in each iteration. The $\mathbb{IKG}$ tends to be proportional to the rate of error reduction, which supports its use as a convergence criteria. The error decreases quickly over the first 25 or so iterations, then slows down as the $\mathbb{IKG}$ becomes small.

Although Bayesian parameterized optimization is able to significantly reduce the value function RMSE from an initial value of 2,200 lb, it does not eliminate it and after the last case the RMSE is still 600 lb. By this point progress has stalled and there does not appear to be much benefit from continuing to add cases with this method. The average value for the structural weight in the control gradient-based optimization wingboxes is around 9,700 lb, so the initial surrogate value function had a relative error of 22% and the final surrogate value function had a relative error of 6%. This was a good accuracy improvement, but considering that the total benefit from using tow-steered composites might be a weight reduction of 5% or less, it does not seem sufficient to give a trustworthy technology benefit estimate.

4.5.4 Another Attempt

Changes

Based on observations from the previous attempt, and to try to bring the error in the surrogate value function down, several modifications were made and Bayesian parameterized

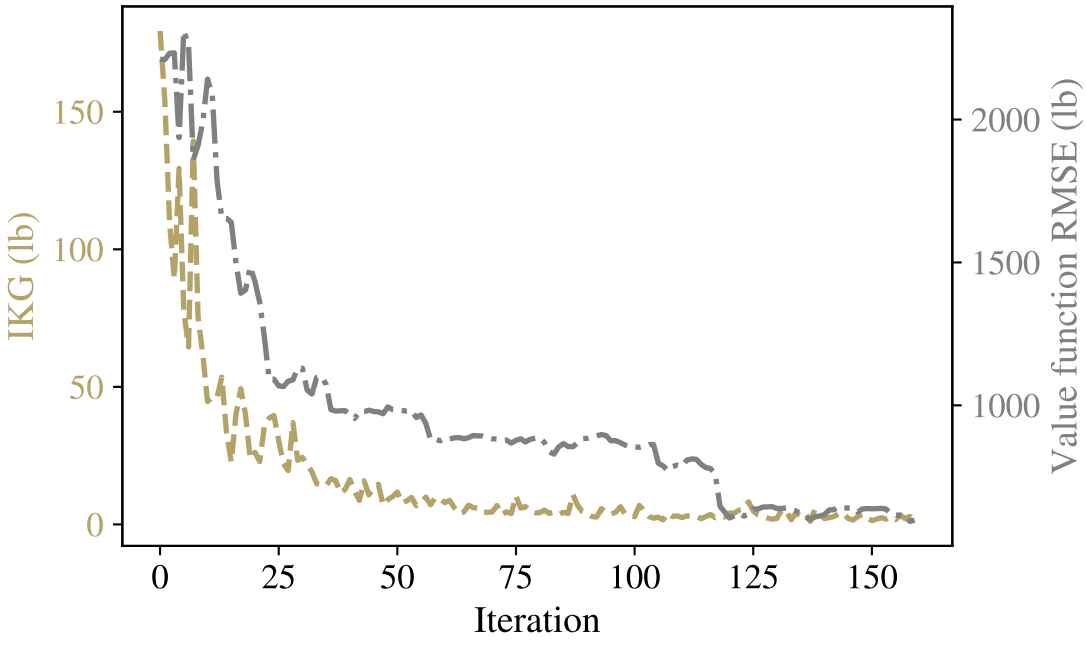

Figure 4.22: Evolution of IKG and surrogate value function error

optimization was reattempted.

First, the results from section 4.4 showed that the objective function's total sensitivity was one of the most important factors affecting the surrogate value function's accuracy. The total sensitivity depends both on the function's behavior and the size of the design space. Since this experiment's objective function is based on the physics of the problem, the remaining way to decrease the total sensitivity is to shrink the design space. Therefore, the range for the choice variables was shrunk from ±60° to ±45°.

The initial data collection was repeated within these new bounds, and the different model form alternatives from table 4.5 were reevaluated. As before, factor analysis directions had the strongest effect and frequently led to inaccurate and overconfident models. The trends for the other options were different (but like last time, relatively weak) and this time the selected model had a quadratic mean, a Matérn ⅝ covariance, and used REML to train the hyperparameters.

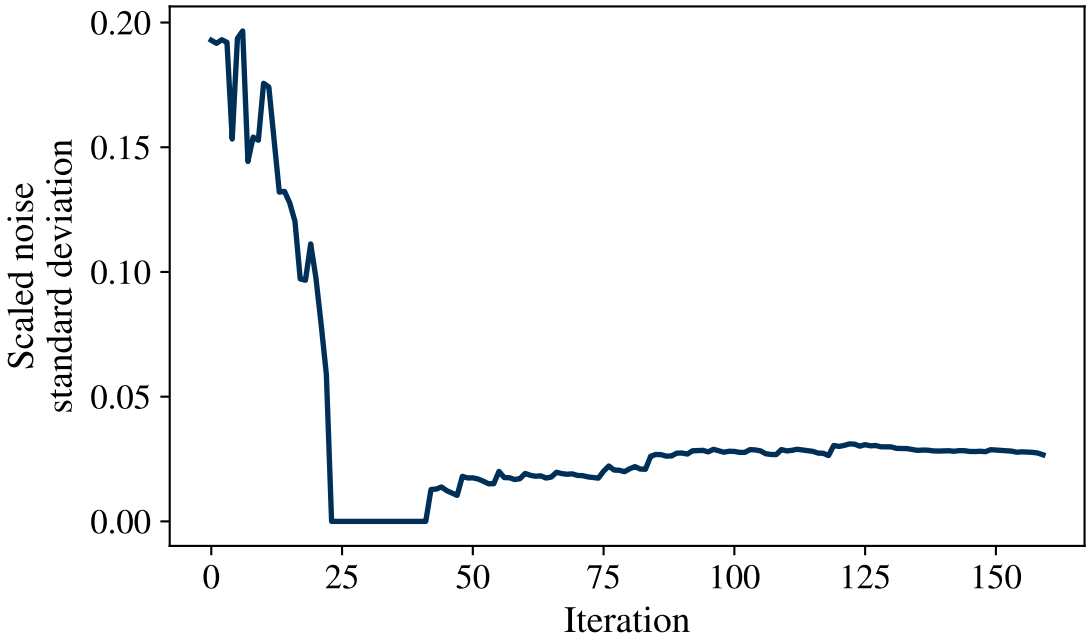

Figure 4.23: Evolution of noise hyperparameter estimate in first attempt

The second observation acted on was how large the initial estimates for the noise hyperparameter were. Figure 4.23 shows how the surrogate's noise estimate progressed through the adaptive iterations. In the first iteration the noise estimate was 685 lb. Although the rate hyperparameters are a bit more esoteric, the noise has a concrete meaning for the physics-based testbed: it represents how much the weight can suddenly jump due to the number of ribs changing, or the numerical tolerance for the weight in the lower-level optimization task. Neither of those effects should cause such a large discrepancy as the estimate indicates. The high noise is also coupled to lower estimates for the rate hyperparameters, with the overall effect being that the model interpreted the data as coming from a noisier and more gradually varying function rather than a less noisy and more quickly varying function. However, as more data was added the noise estimate was revised and by the end it had shrunk to 95 lb.

To investigate this further, the second attempt's model was fit over a sweep of prescribed

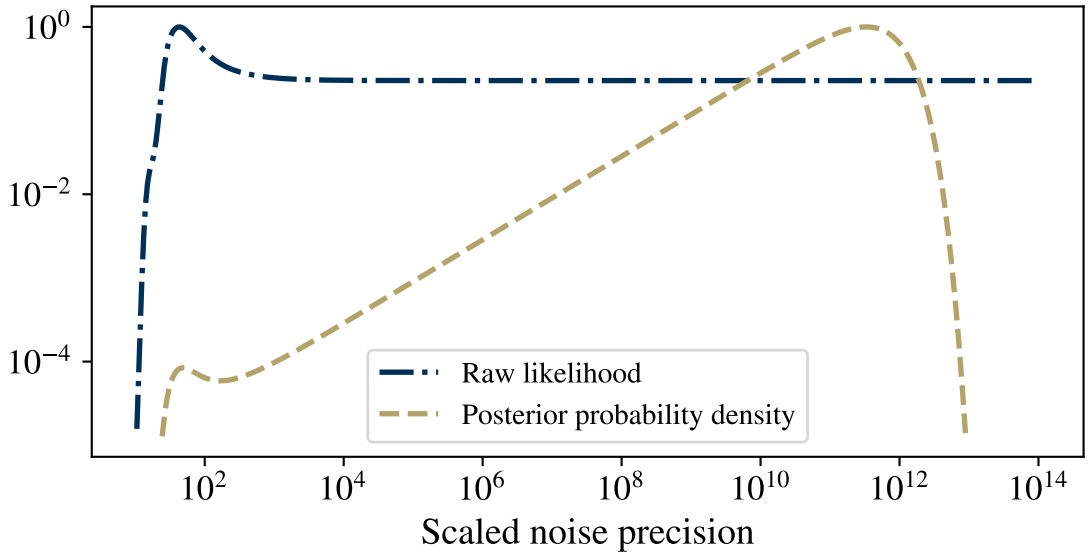

Figure 4.24: Second attempt initial likelihood and posterior conditioned on noise. The blue line shows the likelihood of the estimated hyperparameters with the given noise precision and the gold line shows the posterior probability density of the estimated hyperparameters when the noise precision is given a $\Gamma(\alpha = 1.5,\ \beta = 1.5 \times 10^{-12})$ prior.

noises to observe the log likelihood conditioned on the noise. The blue line in figure 4.24 shows the results, with the noise parameterized by its precision, or inverse variance. The peak likelihood occurs at a low scaled precision of 44, which corresponds to a high standard deviation of 410 lb. However, as the precision increases (variance decreases) the likelihood does *not* continue to decrease and instead forms a ridge of constant likelihood. This might be because the initial stage of data collection used a space-filling design of experiments where the input locations are spaced far apart from each other. This is effective for learning the long- and medium-term behavior across the design space, but does not provide much information about the size of short-term noisy jumps.

Gelman [110] calls such a (hyper)parameter with flatness in its likelihood an underidentified parameter. When this occurs, the peak of the likelihood curve is not necessarily representative of the entire shape. This is an issue for MLE and REML, which are predicated on providing a point estimate for an uncertain value. In a Bayesian context, underidenti-

fication can be dealt with by assigning a prior to the hyperparameter and marginalizing it out. As an intermediate measure, to avoid the expense of a fully Bayesian treatment of the hyperparameters, it was decided to use maximum a posteriori estimation where the likelihood is combined with the prior but still condensed into a point estimate.

The second attempt model's noise precision was given a gamma prior with a shape parameter of 1.5 and a rate parameter of 1.5×10^{-12}. This led to the posterior probability density displayed by the gold line in figure 4.24. The effect of this prior is to move the mode to a high precision so that the initial surrogate model assumes very little noise. However, this prior distribution is relatively wide/weak so as more data is added the likelihood will be able to overcome the prior and produce a different noise estimate. The effect of this prior can be thought of as breaking ties, which occur because the first stage data provide little information about the noise, in favor of the low-noise explanation. When the data do provide information about the noise level there will be no tie and the prior will become irrelevant.

Results

Since the model form included a quadratic mean function, the hyperparameter estimates from the final iteration provide richer information about the objective function than the previous attempt. The coefficients for the linear and quadratic terms of the mean function characterize the global behavior of the objective function, while the covariance sensitivities account for interaction terms (recall the mean only uses pure quadratic terms) and mid-scale nonlinear fluctuations.

Figures 4.25 and 4.26 show the linear and quadratic coefficients, respectively. The linear coefficients are high for the planform parameters, indicating that the weight increases strongly across the design ranges of planform area and especially of aspect ratio. The linear coefficients for the tow orientations are much smaller. The quadratic coefficients are large and positive for the inboard tow paths, and combined with the small linear coefficients this suggests that the optimal inboard tow orientations will be near the center of the design space.

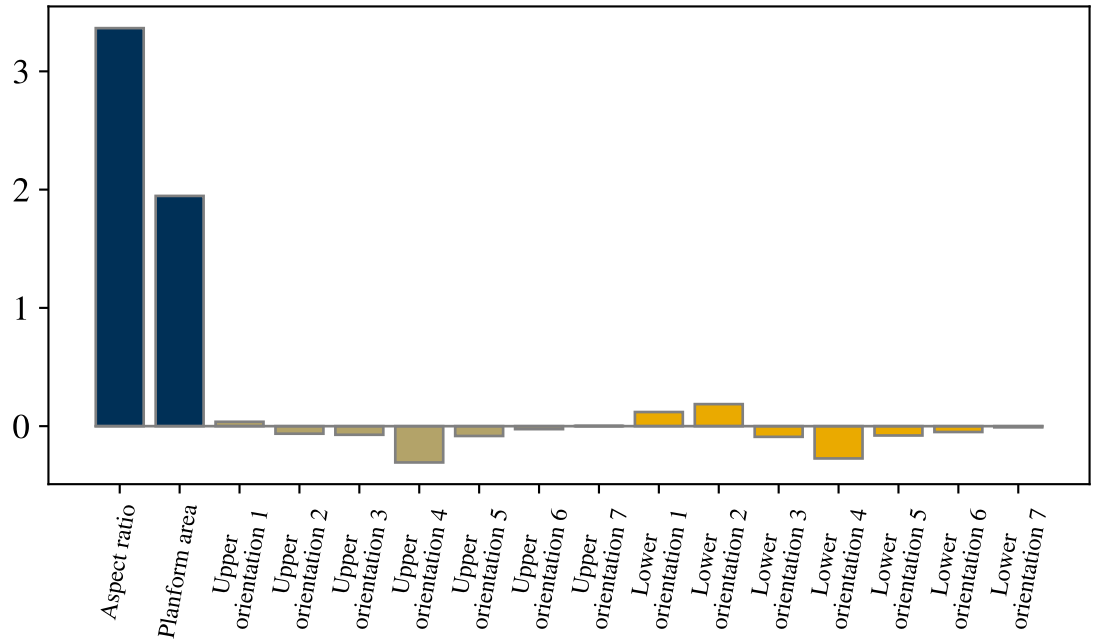

Figure 4.25: Centered and scaled linear coefficients of mean function

There is also some global curvature for the planform parameters which is negative in the case of planform area.

The covariance function's sensitivities in figure 4.27 have similar sizes as the quadratic coefficients, with the most nonlinearity being associated with the inboard tow orientations and some also being associated with the planform parameters. Since the wingtip tow orientations have small linear terms, quadratic terms, and covariance sensitivities, they appear to have no important effect on the weight. Lower coefficients 5 and 6 have no sensitivity but have quadratic terms so they only have a steady curvature effect on the weight.

In this attempt, the initial surrogate value function had an RMSE of 690 lb and the final surrogate one of 200 lb. This was an improvement on the previous attempt, though as shown in figure 4.28 the progress was much more uneven this time. From about iteration 100 onward the error oscillated between an upper and lower level, and did not appreciably decrease. The

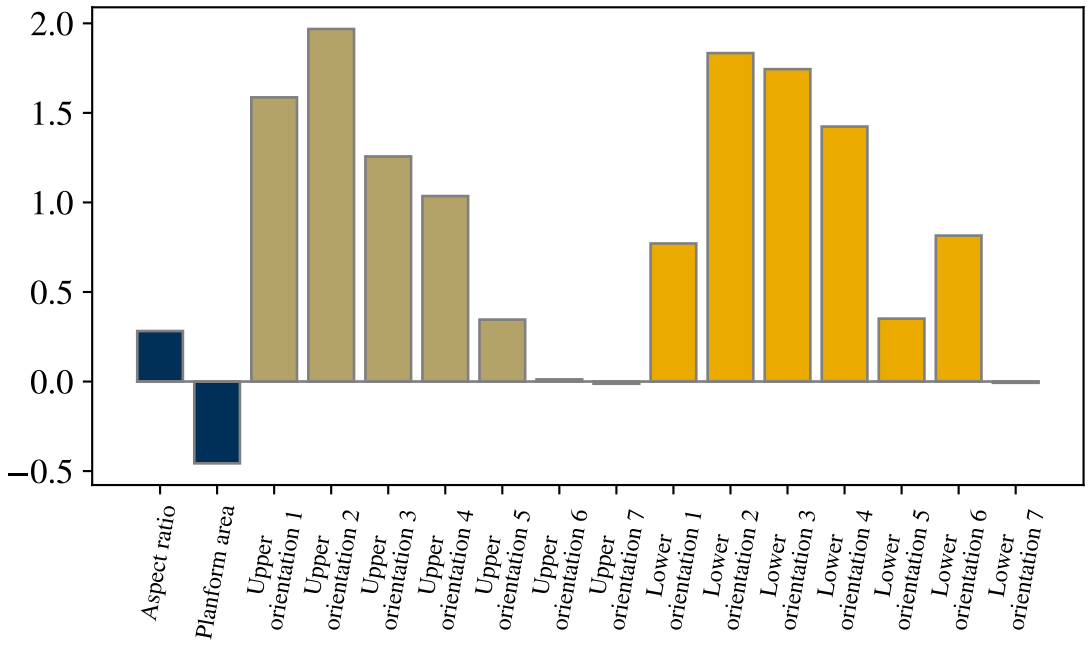

Figure 4.26: Scaled quadratic coefficients of mean function

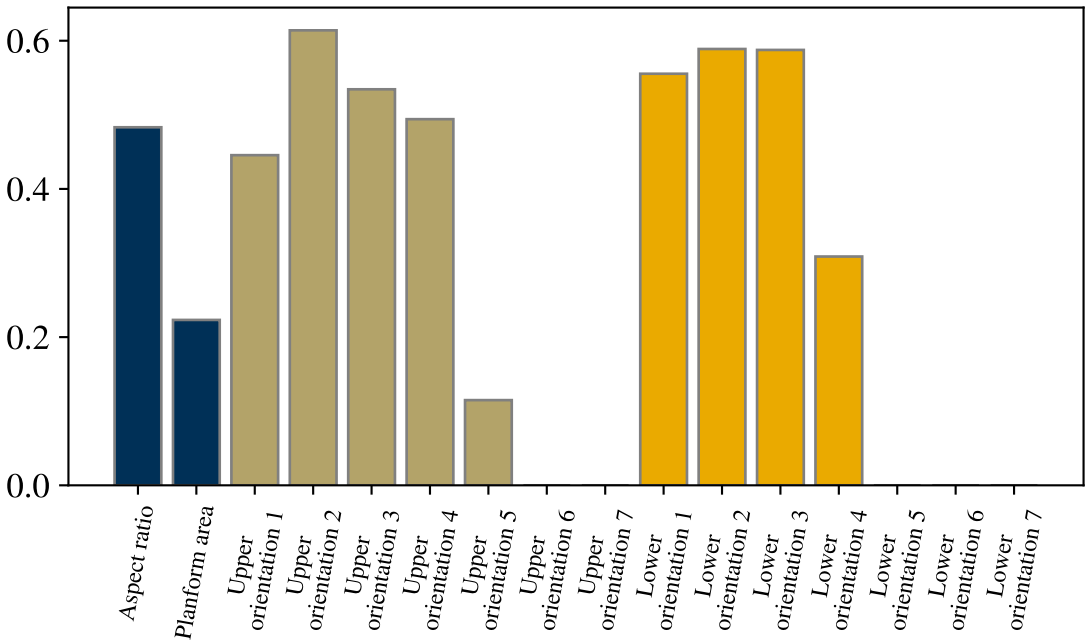

Figure 4.27: Scaled covariance sensitivity hyperparameters

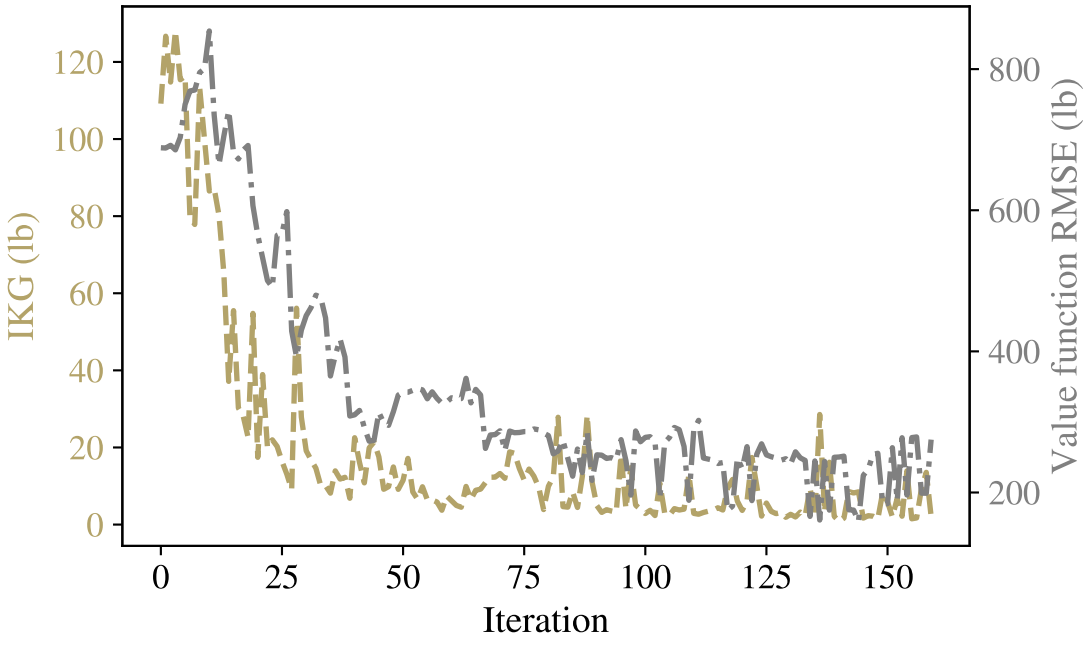

Figure 4.28: Evolution of IKG and surrogate value function error in second attempt

left side of figure 4.29 shows how, after initially estimating practically no noise, the estimated scaled noise started alternating between none and around 4% (corresponding to around 100 lb). The right side shows that the oscillations of the error and the noise estimate fell hand in hand, with the lower error corresponding to iterations where the surrogate had no noise. The alternating behavior is explained by the bimodal shaped of the posterior in figure 4.24; after a certain amount of data had been added the two peaks were at similar heights and exactly which was higher depended on the specific data available in each iteration. This suggests that a stronger prior assumption on the noise level could have worked even better. The minimum error throughout the entire history was 160 lb and occurred in iteration 145.

Figure 4.30 shows plots comparing the surrogate value function with the directly optimized value function. The left side shows that, assuming the directly optimized designs are the true optima, the initial surrogate value function tended to underestimate the optimal weight in lower-weight regions of the parameter space and overestimate the optimal weight

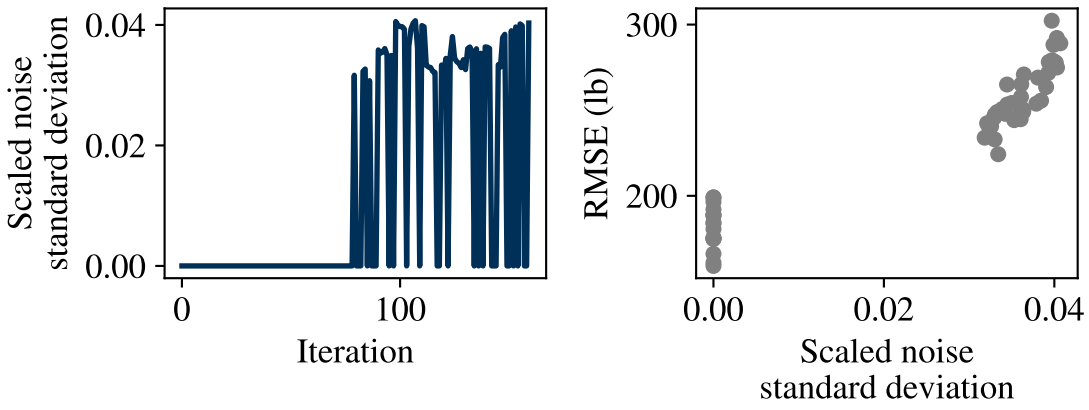

Figure 4.29: Evolution of noise estimate in second attempt, and relationship with RMSE. The left pane shows the noise estimate used in every iteration and the right pane shows a scatterplot of the RMSE against the noise estimate *only for iterations 100 and on*.

in higher-weight regions of the parameter space. There is also a noticeable amount of variability around this trend. On the right side, after the adaptively collected data were added, the two methods produce nearly the same results. The largest discrepancy is for two particular points at the upper range of optimal weights. Whereas the RMSE for the surrogate value function with these points is 200 lb, if these two points are ignored the RMSE drops to 160 lb.

The directly optimized wingboxes are not known to be global minimum weight designs. They were designed by a gradient-based algorithm initialized with a straight spanwise-tow wingbox, which is only able to find the local minimum corresponding to the basin of attraction it starts in. If there were a better local minimum it would not have been found. In contrast, the surrogate-based approach should have been able to approximate the global minimum no matter where in the design space it actually was. The right side of figure 4.30 suggests that Bayesian parameterized optimization might have found superior designs (corresponding to the dots slightly above and to the left of the line), though the amount of possible improvement is very small. To investigate the possibility that Bayesian parameterized optimization found better local minima, the surrogate model's minima were

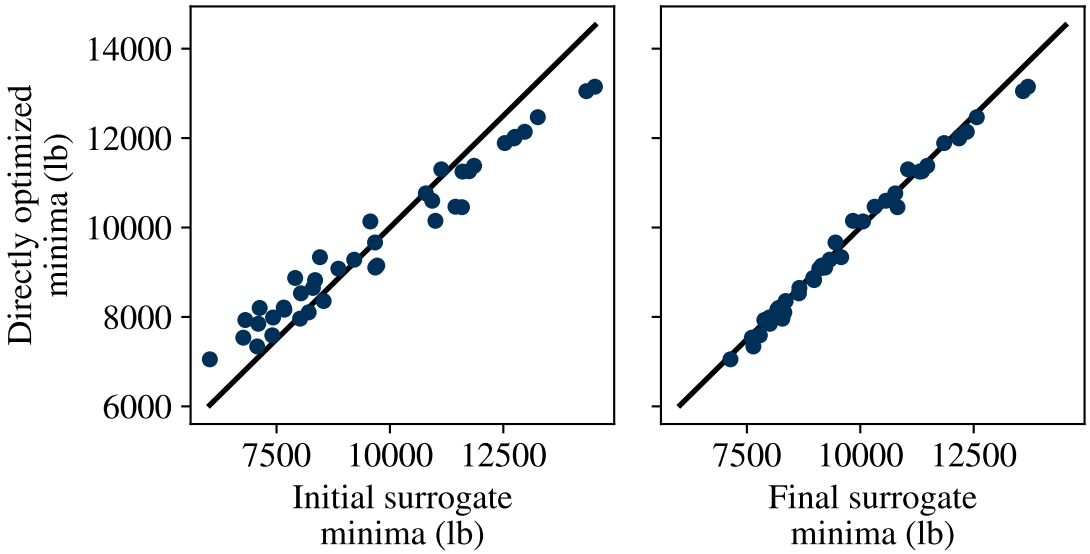

Figure 4.30: Comparison of directly optimized and surrogate-optimized minima

used as starting points for another round of direct gradient-based optimization. The results are shown in figure 4.31. The new designs had practically identical weights as the original directly optimized designs, though they did have different wingtip tow orientations that tended to be more in the extremes of the design space. As mentioned earlier, the surrogate's hyperparameter values suggest that the wingtip tow orientations do not affect the weight. This suggests that there are not multiple local minima with different performances in the tow-steered composite design space, and also that wingtip tow orientations are not important to the weight of the structure.

4.6 Conclusions

Bayesian parameterized optimization, a method for efficiently sampling data to construct a surrogate model of a parameterized optimization problem's objective function, was developed and tested in a couple of experiments. This was motivated by a desire to improve the trade-off between technology benefit and computational cost when creating a surrogate model as part of a process to quantify the impact of using tow-steered composites. The

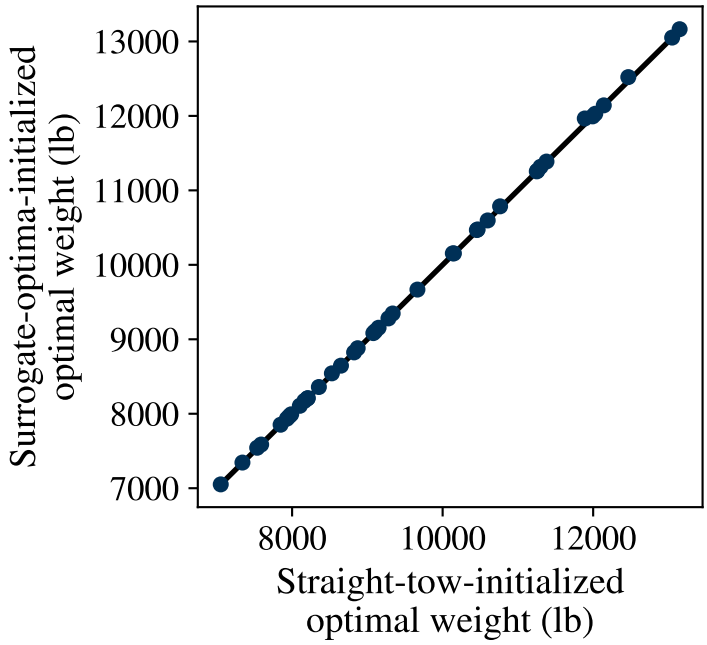

Figure 4.31: Comparison of directly optimized wingboxes with different start points

results of experiment 2 show that the resulting surrogate's accuracy is highly dependent on the characteristics of the objective function, but also that the new method provides an advantage compared to standard space-filling designs of experiments. This advantage increases when there is a large amount of choice variables and choice variable sensitivity, which nicely fits the characteristics of the motivating problem.

In experiment 3 Bayesian parameterized optimization was applied to a more relevant problem that used physics-based structural sizing as the objective function. The first trial of the method did not perform as well as hoped, but this was improved in the second trial by reducing the design ranges. The results suggest that the optimal tow path coefficients tend to be in the middle of the design space anyway, so this compromise does not appear to reduce the structure's performance.

The results of these two experiments support hypothesis 2. In particular, experiment 3 demonstrated that the method can be used to create an accurate surrogate model of a tow-steered composite wingbox's structural weight. However, as mentioned in section 3.3.1, the

testbed structure is somewhat unrealistic because its panels are unstiffened and it is not sized to resist stability failures. Nonetheless, it should have similar overall behavior as a more realistic structural sizing model. The example use case in chapter 6 will provide additional testing of hypothesis 2 by applying Bayesian parameterized optimization to a more realistic structural sizing model.

The experiments showed that the number of variables and their ranges strongly affect the accuracy of the surrogate model, but using this method allows a larger design space to be considered compared to what would otherwise be possible. Since the best tow path coefficients tended to be in the middle of the design space, that suggests that the right way to take advantage of this is to increase the size of the tow path parameterization, rather than increasing the ranges.

Experiment 3 also showed how it is possible to learn characteristics of the physics-based objective function and express them through the surrogate model's hyperparameters. An additional finding was no indication of multiple local minima in the design space. However, because of the simplicity of the testbed that might not be generalizable, and there might be multiple local minima for a more realistic wingbox geometry or for more richly parameterized tow paths.

CHAPTER 5
METHODOLOGY

5.1 Review of Research

Motivation and Goal

At a high level, this research was motivated by a desire to design a new generation of aircraft that are more efficient and have smaller adverse environmental impacts than current aircraft. Tow-steered composites have been identified as a technology that could help play a role in achieving these goals. In order to evaluate the extent to which that is true, and to understand how including the technology transforms the overall vehicle design, the effect of using tow-steered composites needs to be integrated into the system design process.

As a structural technology, the effect of tow-steered composites can be represented in a system design tool through a weight adjustment factor. An appropriate weight adjustment factor can be calculated with physics-based structural sizing. However, because the technology impact could be coupled to the vehicle configuration, and because conceptual design is a fluid process and the final configuration is uncertain, to best support the conceptual design process the weight adjustment should be evaluated across the entire range of likely vehicle configurations. This reasoning led to the research objective:

> Develop a methodology to evaluate the wingbox structural weight reduction benefit provided by tow-steered composite technology across a range of planform configurations

Previous research on tow-steered composite structures has described the designs through high-dimensional parameterizations, and used specialized analysis to efficiently calculate derivatives so that gradient-based optimization algorithms can quickly search the high-

dimensional space. Although successful, this kind of approach places additional requirements on the physics-based model that can increase model development time and make it more challenging to leverage existing models. It is also possible, if the tow-steered composite design space has multiple local minima, that a gradient-based optimization algorithm has difficulty finding the global optimum. In order to complement previous tow-steered composite research and to differentiate the methodology, the methodology had the subgoals of 1) being able to leverage software and existing models that do not provide efficient gradient calculations and 2) to use a global optimization approach to designing the tow paths.

Dimensionality Reduction for Structural Design Problems

Since the physics-based analysis is computationally expensive, a surrogate modeling approach seems like a natural choice for the methodology. If the physics-based analysis is not set up to provide derivatives, the surrogate model can compensate by being cheap to evaluate. However, existing approaches used high-dimensional parameterizations to maximize the benefit from tow steering, and the curse of dimensionality makes it difficult to create an accurate surrogate for a high-dimensional function, leading to gap 1:

> Existing approaches to tow-steered composite design need to use a high-dimensional tow path parameterization to maximize the technology benefit. However, this precludes surrogate modeling, requires gradient-enabled analysis, and makes it difficult to locate the global optimum.

This challenge led to the creation of research question 1:

> How should the tow-steered composite design problem be formulated and the tow paths be parameterized in order to facilitate surrogate modeling, while preserving as much of the technology's benefit as possible?

A review of literature found methods for reduced-order structural design and hierarchical optimization but none of them seemed completely satisfactory for the methodology. Decomposing the design variables into a limited number of upper-level global variables, and a lower level of local variables set by fully stressed design, was determined to be a viable approach for surrogate modeling, but would produce suboptimal designs.

In an attempt to improve this situation, an alternate method based on active subspaces was developed. It was thought that this could be successful because of the success of active subspaces on other nonlinear aerospace engineering problems. Even though derivatives are needed to find an active subspace, if one were found it could be possible to leverage the shape of the active subspace on models that do not provide derivatives as well. However, unlike previous uses of active subspaces, this application involved a constrained optimization problem. A technique for aggregating the objective and constraint functions with an augmented Lagrangian penalty function, and finding the active subspace of said penalty function was developed, and the belief about this method successfully addressing research question 1 was stated in hypothesis 1:

> If the objective and constraint functions of the tow-steered composite design problem are aggregated by an augmented Lagrangian penalty function, and if the active subspace of this penalty function is found, then that active subspace can be used as a low-dimensional parameterization of the problem suitable for surrogate modeling without sacrificing much performance. This is because the penalty function's minimum is the solution of the original problem and because the active subspace approach is able to find a reduced set of directions that explain most of a function's behavior.

Because this method was novel, an experiment was needed to demonstrate its effectiveness and to test the hypothesis. There were also several parameters of the method that were suspected to affect its success, but for which the best settings were unknown. Experiment 1 was conducted to address these needs, and used a simplified testbed that included the most

important physics to the tow-steered composite design problem. The results generated from the experiment showed that the method's parameters did affect its success and did produce clear trends, but also showed that the method was unable to deliver the desired results. Three important design directions were clearly identified and had rational explanations, but they did not account for enough of the function's behavior and the subsequent directions were too similar in importance to yield the desired dimensionality reduction or insight.

Therefore, the results of the experiment did not support hypothesis 1. As a result, the methodology will call for the use of the multilevel formulation with a surrogate model of the top-level global design variables and fully stressed design type approach for the lower level local structural dimensions. This approach enables the use of surrogate modeling but imposes some important limitations on the methodology: 1) the technology benefit that the methodology can measure will be reduced due to the coarse parameterization of the tow orientations, 2) the structural designs generated by the methodology will be good but not optimal due to the use of a fully stressed design type approach.

Surrogate Modeling of Parameterized Optimization Problems

The other research focus was on how to collect data for the surrogate model. Increasing the dimensionality of the tow path parameterization improves the technology benefit but makes surrogate modeling more difficult. This drives a need for the data collection to be as efficient as possible, as stated in gap 2:

> Because the trade-off between technology benefit and computational cost is significant, it must be made as efficient as possible. This creates the need to evaluate the expensive physics-based analysis at carefully selected points that reflect the system's behavior and the experimental goals.

Designs of experiments are used to efficiently plan out data collection. Standard designs of experiments for computer simulations try to uniformly sample the design space and are

suitable for creating a surrogate model that is accurate over the entire space. However, it was observed that the surrogate does not actually need to be accurate over the entire space for the purpose of finding the best tow path design for each planform configuration. Therefore, the sampling approach should prioritize regions of the design space that help achieve this goal, leading to research question 2:

> What design of experiments should be used to sample data from the expensive physics-based analysis in order to efficiently make a surrogate model that can be used to accurately determine the weight of the optimal tow path structure for all possible planform settings?

This goal was classified as a parameterized optimization problem: the surrogate must be accurate across all the planform settings to provide maximum freedom/information for vehicle design, but the surrogate only needs to be accurate in regions of the tow path space that are important for finding the optimum.

There are many surrogate-based optimization methods in literature, and similar surrogate-based methods designed for other kinds of problems. However, none were found to be specifically devised for parameterized optimization problems. One category of these methods, Bayesian optimization, was found to provide a useful framework for addressing new problem types by expressing the surrogate's prediction probabilistically and using an acquisition function that embodies the problem's goal in order to guide the selection of new data points.

Therefore, Bayesian parameterized optimization was developed as a method to efficiently collect data for the desired surrogate model, and the integrated knowledge gradient acquisition function was developed to drive the selection of new points. The reasoning behind the creation of this method was embodied in hypothesis 2:

> If Bayesian parameterized optimization is used to drive data collection for the tow-steered composite design problem, then the data will be collected efficiently

> and the resulting surrogate model will be able to recover accurate solutions of the parameterized optimization problem, because Bayesian optimization methods are able to focus the improvement of surrogate model accuracy in regions of the design space that are likely to contain solutions.

This hypothesis was tested over the course of two experiments. Experiment 2 applied the method to a range of artificially generated functions with different characteristics in order to determine if the method provides an advantage relative to a space-filling design of experiments, and how the characteristics of the objective function affect the surrogate's accuracy. The results of this experiment showed that Bayesian parameterized optimization does provide an advantage, particularly with increasing dimensionality and an abstract quantity representing the objective function's total nonlinearity. The experiment also found a relatively smaller benefit for objective functions with more choice variables than parameters, which is the case for the tow-steered composite design problem if there are just a few planform variables and many tow orientation variables.

Experiment 2 lent some support to the hypothesis but because it used artificial test functions the method's ability to succeed on the intended application was still in question. Experiment 3 was performed to give more confidence that Bayesian parameterized optimization works on tow-steered composite design problems by using the method to create a surrogate model of the same simple physics-based testbed used in experiment 1. The first attempt showed some success but also had somewhat low accuracy. In accordance with the result from experiment 2 that the objective function's total sensitivity was the most important factor affecting accuracy, the size of the design space was modestly reduced and the method was retried. This second attempt showed a good improvement in accuracy, suggested that there likely are not multiple local minima associated with the coarse tow path parameterization, and provided insight about the tow-steered design problem's objective function from the surrogate model's hyperparameters. Together, experiments 2 and 3 provide complementary support for hypothesis 2, with the caveat that the size of the design space has

a strong effect on the surrogate model's accuracy, and Bayesian parameterized optimization will therefore be included in the methodology.

5.2 Methodology Description

Based on the knowledge gained from the previously described research and experiments, the following methodology is proposed for estimating the structural weight reduction benefit provided by tow-steered composite technology across a range of planform configurations. The technology impacts obtained by following this methodology can then be inserted into a system-level design model in order to propagate the effect from the technology to the system level and across disciplines.

This methodology builds on previous work for structural technology performance estimation by providing an efficient sampling method for examining the entirety of the vehicle configuration space and the favorable regions of the structural design space. The methodology also distinguishes itself from other tow-steered composite wingbox design methods by not requiring the use of gradient-enabled analysis software, which makes it easier to leverage existing models lacking that feature, and by having the ability to search for the globally optimal tow paths. The steps of the methodology are:

METHODOLOGY

1. Formulate problem
2. Develop physics-based modeling environment
3. Run initial stage of experiment
4. Select and validate surrogate model form
5. If surrogate model accuracy needs improvement, run sequential stages of experiment
6. Validate surrogate value function

The product obtained by following the methodology is a surrogate model that is cheap to evaluate and can be optimized quickly to evaluate the weight reduction from tow-steered

composite technology at any point in the planform design space. More details on the individual steps of the methodology follow.

Step 1: Formulate Problem

Formulating the problem consists of selecting the parameters and choice variables of the parameterized optimization problem, as well as their experimental ranges, and also identifying attributes of the vehicle configuration and structure that will be represented in the structural model.

The parameter are intended to be major wing planform variables like the area and aspect ratio, and could be extended to include sweep, taper ratio, and twist if these quantities were thought to affect the technology benefit. It would also be possible to include other system-level variables separate from the planform such as takeoff gross weight and engine weight. However, each added variable and its range needs to be balanced against the increased difficulty of making an accurate surrogate model in higher dimensions.

The choice variables consist of the structural design variables that have a global effect and are not particularly relevant to system-level design. This most importantly includes the tow paths but could also include elements of the structural layout like rib or stiffener spacing, or spar positions. It could also include composite laminate variables like the percentage of zero degree plies. Like the parameters, the number of choice variables needs to be limited so that the total number of variables is about twenty at most. The results from experiment 3 indicate that a range of ±60° for the tow orientations might be too wide to make an accurate surrogate model, and that reducing the range improves accuracy.

Other information needed to fully define the problem includes the system-level variables that will be treated as constants and the load cases that will be used to size the structure. The local structural design problems for each part of the structure must also be formulated by defining the regions that they apply to, the structural material properties, the stiffened panel shapes, and the design parameterization of the structural dimensions. There will be

too many local structural variables for a surrogate model so they will be implicitly included by designing them with a fully stressed type approach based on the parameter and choice variable settings.

Step 2: Develop Physics-Based Modeling Environment

Physics-based analysis is needed to map a chosen parameter and choice variable setting to structural weight. This starts with a model of the configuration's outer mold line geometry. To perform structural mechanics analysis, a finite element model must be created by generating geometry for the skins, ribs, and spars within the outer mold line, meshing that geometry, and assigning properties to the elements. Rather than explicitly modeling the details of stiffened panels, a homogenized representation can be used and gives a good trade-off of accuracy and expense. In this case the stiffened panel dimensions are reflected by the stiffness properties of the finite elements.

The model must also represent the important external load sources. These include inertial forces from the structure's own mass, the engines, and fuel, as well as the aerodynamic forces. Vortex lattice aerodynamic models provide a good trade-off of accuracy and expense for this kind of structural weight estimation procedure. Because aeroelastic interaction is important to the performance of tow-steered composites, the aerodynamic model should also be connected to the structural model using a mechanism such as a spline. This will transfer displacements from the structural model to the aerodynamic model, and forces from the aerodynamic model to the structural model.

Finally, the fully stressed design of the local structural variables must be implemented. The internal forces calculated by the aeroelastic analysis will be used in structural failure theories to evaluate the margin of safety for each component, and local structural optimization will be performed on each component to minimize weight with a satisfactory margin. Commonly used types of failure theories are material strength, buckling of global panels, buckling of local panel objects, and crippling. The overall procedure must be set up to

iterate between calculating the internal forces and updating the local structural designs until the two converge, since they are coupled.

Over the course of executing the methodology the objective function will be called hundreds of times and, since the effect of the planform is an emphasis, almost every one of those calls will analyze a unique geometry. Manually constructing all these models would be totally impractical; the methodology requires an automated capability to vary the outer mold line geometry, layout the structural geometry within the wing, create the coupled meshes of the structural and aerodynamic geometries, and assign properties to the mesh elements.

Step 3: Run Initial Stage of Experiment

The initial stage of data collection should plan out a relatively large number of cases in advance; a standard rule of thumb is $10d$ cases. This will enable the initial surrogate model to have a good understanding of the coarse trends across each dimension even if it is inaccurate, and thereby will help with the selection of additional points. The results of experiment 3 suggest that there is nonlinearity beyond simple curvature associated with all the variables except tow orientations at the wingtip, so a space-filling design of experiments is appropriate. A maximum entropy design [130] could take advantage of these differences between the dimensions, but a maximin Latin hypercube was shown to perform fine.

Step 4: Select and Validate Surrogate Model Form

The methodology requires that the surrogate model be a Bayesian regression model so that it can provide probabilistic predictions of objective function values, but makes no demands on the exact form of the model. The Gaussian processes with stationary covariance functions used in experiment 3 performed well. It might be possible to get better predictions by using a more flexible model but, as seen with the factor analysis directions, more flexibility comes with a risk of overfitting that must be guarded against.

The actual by predicted plot and the normal quantile plot are good for detecting bad

model behavior. However, leave-one-out cross-validation might not be able to detect bad behavior from a more flexible model form, so if a more flexible model form is being considered then it is important to have an independent set of validation data.

Step 5: Run Sequential Stages of Experiment

If the surrogate model is already accurate then there is no need to collect further data. However, the results of experiment 3 suggest that this is unlikely unless the problem was formulated with a very narrow design space. In that case, there is an opportunity to go back to step 1 and reformulate the problem to consider more variables or wider design ranges.

If the surrogate model is not accurate enough, then Bayesian parameterized optimization with the integrated knowledge gradient acquisition function should be used to sequentially add data to improve the surrogate. The integrated knowledge gradient can be approximately calculated using the sampling approach from equation (4.17), in which case the variance reduction strategies of antithetic normal samples and stratified sampling should be used. Regardless of how $\mathbb{IKG}$ is calculated, because it is highly multimodal a global optimization strategy such as multistart optimization should be used to maximize it.

This approach can be used either with a fixed budget of additional cases or using the maximal $\mathbb{IKG}$ as a termination criterion. In experiment 3, the maximal $\mathbb{IKG}$ started at around ~1% of the typical structural weight in the first iteration. The surrogate model's improvement started slowing significantly after this value fell to around ~0.05% of the typical structural weight, which provides a reasonable setting for the termination threshold. However, $\mathbb{IKG}$ did not decrease monotonically throughout the iterations, so to avoid premature termination one could wait until it falls below the threshold value for several consecutive iterations.

Step 6: Validate Surrogate Value Function

After collecting all the data, as a final check, the surrogate model should be optimized at multiple parameter settings, and the predicted minima should be evaluated with the physics-

based analysis. The integrated knowledge gradient favors samples that provide the most information about the optima, and not necessarily samples that are themselves optimal with respect to the choice variables. It is therefore prudent to sample some predicted optima to make sure that the surrogate model is correct and to quantify its accuracy.

If the physics-based model does provide efficiently calculated derivatives, then there is no good way to directly evaluate the value function, and as a result there is no good way to directly validate the surrogate value function. However, by evaluating these additional cases it is at least possible to validate the surrogate objective function's accuracy in the locations that are thought to contain the optima. If the surrogate objective function is accurate in the regions containing the true optima, then the surrogate value function would also be accurate.

5.2.1 Thesis Statement

OVERARCHING HYPOTHESIS

The proposed methodology realizes the research objective by providing the capability to evaluate the wingbox structural weight reduction benefit granted by tow-steered composite technology across a range of planform configurations. Its surrogate-based approach naturally lends itself to finding globally optimal tow path designs, and though this also limits the dimensionality, an efficient sampling technique helps to maximize the technology benefit subject to this limitation. Further, the methodology promotes quick turnaround by removing the requirement that the physics-based analysis provide gradients.

CHAPTER 6
EXAMPLE USE CASE

Providing a full empirical validation of the overarching hypothesis is beyond the scope of this work. To do that would require applying the methodology to a comprehensive set of vehicle concepts and benchmarking its performance against the other relevant techniques in each case. However, the overarching hypothesis is already somewhat justified by the reasoning and experimentation that was carried out while developing it. To add further support to the overarching hypothesis, this chapter showcases the thesis methodology in an example use case to demonstrate that it can be used in a more realistic and detailed setting to solve problems of practical interest.

The vehicle configuration considered in this use case is the Common Research Model (CRM) [78]. This configuration is representative of a conventional wide-body commercial transport and was originally developed to serve as a standard benchmark for computational fluid mechanics methods. A model of the CRM's geometry is shown in figure 6.1 and baseline geometry, weight, and structural layout quantities are listed in table 6.1. Although the CRM has a defined outer mold line, this use case will proceed as if it is the conceptual design stage and a range of planform alternatives are still being considered. The goal of this exercise is to quantify the benefit that tow-steered composites can provide across all the potential planforms.

6.1 Problem Formulation

The problem was formulated as a parameterized optimization problem in which the parameters are the wing's area and aspect ratio, and the choice variables are the coefficients of spanwise B-splines specifying the tow orientations on the upper and lower skin. The orientation splines have seven coefficients each. An additional variable called AEQR was

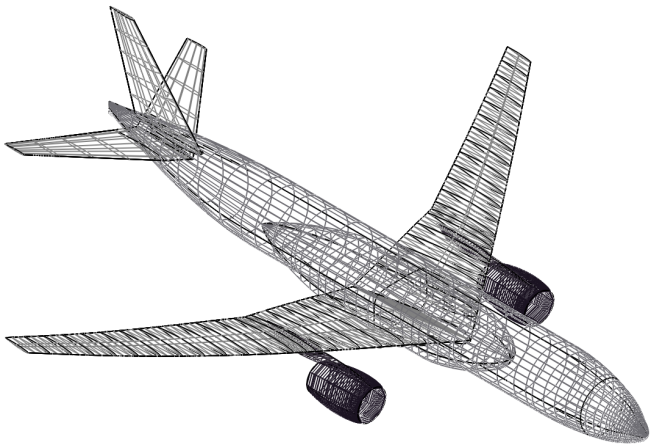

Figure 6.1: Baseline CRM geometry for example use case

Table 6.1: Example use case geometry, weight, and structural layout

Category	**Variable**	**Value**	
Baseline geometry	planform area	4,130	ft^2
	aspect ratio	9	
	span	192.8	ft
	taper ratio	.275	
	quarter-chord sweep	35	deg
Weight	takeoff gross weight	550,000	lb
	individual engine weight	19,000	lb
Structural layout	front spar chord	15	%
	rear spar chord	70	%
	nominal rib spacing	28.8	in

Table 6.2: Variable ranges for example use case

Group	Variable	Settings	
		Low	**High**
Parameters	planform area	3,800 ft^2	5,000 ft^2
	aspect ratio	8	15
	AEQR	0	1
Choice variables	inboard coefficients	-20 deg	20 deg
	outboard coefficients	-45 deg	45 deg

also included in the parameters to control the aeroelasticity of the physics-based model so that the importance of this interaction could be studied. More details on AEQR are given in section 6.2. Including this variable also demonstrates flexibility by the methodology to accommodate parameters related to the analysis itself in addition to the configuration. In total there are 17 input variables split into 3 parameters and 14 choice variables.

The ranges for all the variables are listed in table 6.2. The planform variables are straightforward variations about their baseline values. AEQR's range from 0 to 1 corresponds to a continuous shift from no aeroelasticity to full aeroelasticity in the physics-based model. The tow paths are parameterized by cardinal B-splines with seven coefficients, and with all angles measured relative to the line passing from the rear spar's side-of-body location to its tip location. Because the results of experiment 3 suggested that the optimal tow orientations for the inboard section of the wing are in the interior of the design space, and because wider design ranges increase the difficulty of making an accurate surrogate model, a range of ±20° was used for both splines' four innermost coefficients. The range for the three outermost coefficients was set to ±45°.

Blade stiffeners were chosen for all the structural panels. A cross section of this panel concept and its associated local design variables is shown in figure 6.2. Composite AS4/3502 unitape [131] was selected as the ply material throughout the structure. Its mechanical

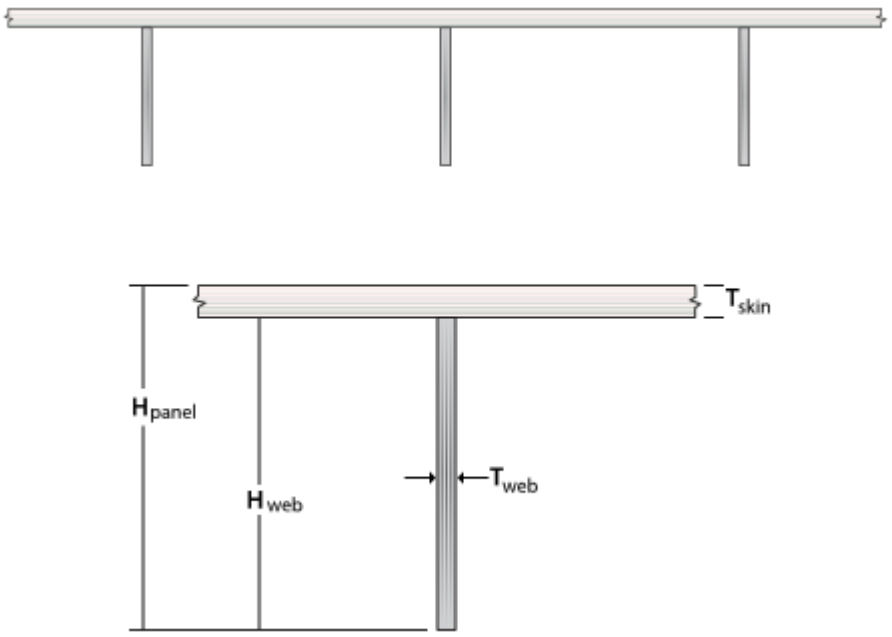

Figure 6.2: Blade-stiffened panel cross section and dimensions. **Credits: Collier Research Corporation** [132]

properties, including strength knockdowns accounting for barely visible impact damage, are displayed in table 6.3. The tow-steered skin panels were given a hard laminate with ply percentages of 62.5% 0°, 25% ±45°, 12.5% 90°, and with properties shown in table 6.4. Note that the large fraction of 0° plies makes this laminate very stiff in the direction of the tow steering. A softer laminate with ply percentages of 20% 0°, 60% ±45°, 20% 90° and properties shown in table 6.5 was picked for the predominately shear-loaded spar and rib panels. For the blade stiffeners, which are under a combined axial and shear load, a laminate with ply percentages of 45% 0°, 45% ±45°, 10% 90° and properties listed in table 6.6 was selected.

Two static maneuver load cases, the pull-up and push-down at maximum limit load factors listed in table 6.7, were chosen to size the structure. For both load cases, each part of the structure must have positive margins of safety for all failure analyses when a safety factor of 1.5 is applied.

Table 6.3: Properties of AS4/3502 tape with strength knockdowns

Attribute	Description	Value
E_1	longitudinal modulus	19.3 Msi
E_2	transverse modulus	1.35 Msi
ν	Poisson ratio	.34
G_{12}	longitudinal-transverse shear modulus	0.543 Msi
G_{13}	lateral-longitudinal shear modulus	0.543 Msi
G_{23}	lateral-transverse shear modulus	0.543 Msi
ρ	density	0.0570 lb/in^3
X_t	longitudinal tensile strength	77.5 ksi
X_c	longitudinal compressive strength	60.0 ksi
Y_t	transverse tensile strength	10.7 ksi
Y_c	transverse compressive strength	10.4 ksi
S	in-plane shear strength	4.8 ksi

Table 6.4: Properties of 62.5% 0°, 25% ±45°, 12.5% 90° AS4/3502 skin laminate

Attribute	Description	Value
E_1	longitudinal modulus	13.3 Msi
E_2	transverse modulus	4.6 Msi
ν	Poisson ratio	.33
G_{12}	longitudinal-transverse shear modulus	1.65 Msi
G_{13}	lateral-longitudinal shear modulus	0.543 Msi
G_{23}	lateral-transverse shear modulus	0.543 Msi
ρ	density	0.0570 lb/in^3
X_t	longitudinal tensile strength	53.4 ksi
X_c	longitudinal compressive strength	41.5 ksi
Y_t	transverse tensile strength	18.4 ksi
Y_c	transverse compressive strength	14.7 ksi
S	in-plane shear strength	10.0 ksi

Table 6.5: Properties of 20% 0°, 60% ±45°, 20% 90° AS4/3502 spar/rib laminate

Attribute	Description	Value
E_1	longitudinal modulus	6.5 Msi
E_2	transverse modulus	6.5 Msi
ν	Poisson ratio	.40
G_{12}	longitudinal-transverse shear modulus	3.11 Msi
G_{13}	lateral-longitudinal shear modulus	0.545 Msi
G_{23}	lateral-transverse shear modulus	0.545 Msi
ρ	density	0.0570 lb/in^3
X_t	longitudinal tensile strength	25.7 ksi
X_c	longitudinal compressive strength	20.2 ksi
Y_t	transverse tensile strength	25.7 ksi
Y_c	transverse compressive strength	20.2 ksi
S	in-plane shear strength	17.5 ksi

Table 6.6: Properties of 45% 0°, 45% ±45°, 10% 90° AS4/3502 stiffener laminate

Attribute	Description	Value
E_1	longitudinal modulus	10.4 Msi
E_2	transverse modulus	4.7 Msi
ν	Poisson ratio	.47
G_{12}	longitudinal-transverse shear modulus	2.46 Msi
G_{13}	lateral-longitudinal shear modulus	0.547 Msi
G_{23}	lateral-transverse shear modulus	0.547 Msi
ρ	density	0.0570 lb/in^3
X_t	longitudinal tensile strength	40.7 ksi
X_c	longitudinal compressive strength	31.7 ksi
Y_t	transverse tensile strength	19.5 ksi
Y_c	transverse compressive strength	15.4 ksi
S	in-plane shear strength	13.9 ksi

Table 6.7: Example use case load cases

Name	Load factor	Fuel level (%)	Altitude (ft)	Mach	Static margin
pull-up	2.5	100	30,000	0.83	0.1
push-down	-1	100	30,000	0.83	0.1

6.2 Physics-Based Modeling Environment

6.2.1 Formulation

A physics-based computational environment was needed to translate all of the information compiled in the previous step—vehicle weight, outer mold line geometry, planform and structural layout parameters, stiffened panel concept, composite ply and laminate properties, and sizing load cases—into an estimated structural weight. This computational environment must fully automate model construction, analysis, and design to allow hundreds of different planform and tow path settings to be evaluated. A similar approach to the physics-based testbed used in experiments 1 and 3 was followed here: the structure's mechanical behavior is represented with a linear elastic shell finite element model, aerodynamic pressure forces are obtained from a vortex lattice model, and a spline interpolates the structural displacements to link the models and enable aeroelastic interaction. Background information about these techniques can be found in section 3.3.1. The finite element model uses a homogenized representation of stiffened panels and does not include detailed features like cutouts and access panels.

However, there are also important differences between this physics-based environment and the one used in the simple physics-based testbed. These differences are the outer mold line and structural geometries, the stiffeners, the consideration of more structural failure modes, the local panel design formulation, and the fact that derivatives are not needed. Except for derivatives not being needed, these differences increase the realism of the use case while also increasing the amount of effort needed to develop the computational

environment.

Outer Mold Line Geometry

The first step in building the model is creating the outer mold line geometry, which must be parametrically defined in order to assess configurations with different planform areas and aspect ratios. Because the wing has sweep, taper, a yehudi break, and realistic airfoil cross sections, parametric variation is more difficult than merely stretching and rescaling the wing. For example, if the aspect ratio were increased by uniformly increasing the wing's span and decreasing its chord, then the yehudi break would move outboard by an unreasonable amount.

The wing's three-dimensional geometry can be represented by a sequence of airfoil cross sections arranged from root to tip. Each cross section has a normalized airfoil shape, chord length, thickness-to-chord ratio, and twist. Interpolating the perimeters of these sections yields a sequence of trapezoidal wing segments with their own span length, sweep, taper, and dihedral. The environment must implement logic that uses the specified high-level planform parameters to modify these sections and segments to produce the correct wing shape implied by the planform parameters.

Structural Geometry

After the outer mold line geometry is set the structural geometry can be generated. The first part of this process is to place surfaces for the front and rear spars, and the side-of-body and tip ribs, as these components form the outline of the main wingbox. It is convenient at this stage to use surfaces that extend well above and below the outer mold line to ensure the parts take up the whole interior space; the sections outside the outer mold line can be trimmed away later. As seen in table 6.1, the spar locations are defined in percent chord of the wing planform, so the environment must be aware of the planform geometry to construct the structural geometry. As an added detail, the rear spar should have a kink at the yehudi

break.

After these surfaces have been generated, rib surfaces can added in even spacings from the side-of-body rib to the tip rib. However, the side-of-body rib is oriented streamwise while these other ribs are yawed with the wing's sweep. This causes the side-of-body rib to converge with another rib, forming a three-way intersection where the side-of-body rib meets the rear spar. Depending on the rib spacing and the chord length, there might also be y-ribs extending from the front spar to a location in the middle of the side-of-body rib (rather than to the rear spar). Finally, to create the center wingbox, which is within the fuselage, the side-of-body rib's cross section can be swept to the centerline with more center rib surfaces added in.

After the initial part surfaces have been created, they must be fused with each other and the upper and lower skin surfaces to ensure proper topological connectivity. This consists of finding all point and edge intersections between the surfaces and ensuring that the same point/edge is used for all the surfaces that contain it (contrasted with incorrectly using duplicate points/edges). This will be important so that mesh elements on adjacent faces are correctly connected. After fusing, sections of the surfaces that extend outside the outer mold line can be identified and removed, and since only the main wingbox is being modeled, sections of the parts that extend beyond the front and rear spars are also removed.

This entire process can be fragile, as its difficult to consider all the ways the shapes might interact with each other as they are created. For example, many shape intersections need to be calculated throughout this process, and depending on numerical tolerances intersections might be detected or not detected when the opposite behavior is desired. There is also the possibility of creating geometry that leads to a bad mesh later. For instance, if one of the ribs almost intersects the edge at the rear spar's kink, then this would result in a very thin face between the rear spar's kink and the rear spar/rib intersection. This thin face would force the local mesh to either be very fine or have elements with poor aspect ratios.

Stiffeners

Stiffened panels add realism to the modeled structure, but also add steps to the analysis procedure both before and after the finite element analysis. In order to get proper shell element stiffness properties, each panel must be homogenized into effective panel stiffnesses using the panel cross section's stiffness-weighted area and moments [31]. Later in the process, after finite element analysis has determined the internal loads in each element, the element loads must be split into separate loads on the panel's skin and stiffeners. Separate skin and stiffener loads are needed to independently determine whether these components fail. These steps are needed because the stiffeners are not explicitly represented in the finite element mesh, but this is still easier than including the stiffeners in the mesh because then the structural geometry and mesh would need to be regenerated every time the panel designs were changed.

Real structural panels use stiffeners because they greatly increase the panel's bending stiffness and consequently its resistance to stability failures like buckling. In contrast to the testbed used in experiments 1 and 3, this environment models stiffened panel structure so it is fair to consider these kinds of failures in addition to material strength. In fact, these kinds of failures must be considered so that there is a physics-based reason to select the relative sizes of each panel's skin and stiffener components.

Another effect of the stiffeners is to increase the number of design variables. As seen in figure 6.2, there are three independent local design variables that describe the shape of the stiffened panel system. The spacing between stiffeners adds a fourth design variable. This means that there are four times as many design variables than if basic unstiffened panels were used. If a stiffener shape with more dimensions (like an "I") were used, there would be that many more design variables.

Panel Design

Since the wing planform and the tow paths will be specified in each execution of the physics-based weight estimation environment, the environment must complete the structural design by determining the best stiffened panel dimensions to arrive at an estimated structural weight. As planned, the large number of panel dimension design variables will be handled by a fully stressed design type approach where the large overall design problem is split into separate local problems, each with just a handful of design variables.

For each individual problem the stiffened panel dimensions are optimized for minimum weight while satisfying the safety margin requirements for the failure analyses, which depend on the finite element loads. As discussed, the panel designs affect the internal load paths, so the finite element model will need to be reanalyzed with updated stiffnesses and iterated with the panel design until the two converge.

6.2.2 Implementation

RADE

As shown in the preceding section, many different functionalities are needed in order to produce a fully automated program for structural weight estimation. The major steps of this process are:

1. Modify the outer mold line geometry according to the planform variables
2. Place structural geometry within the outer mold line
3. Mesh the planform for the vortex lattice model and the structural geometry for the finite element model
4. Assemble the mesh, stiffness properties, etc. into the models
5. Analyze the models to determine the internal loads
6. Design the local panel structure to withstand the internal loads

Fortunately, there are already many third-party software tools that provide the needed

capabilities for individual steps of this process. However, it remains a major task to integrate these components together with new code to form the overall program. Data needs to be processed and transformed at each step along the way so that the third-party tools receive input in their specified formats. These tools also often have the flexibility to solve broad classes of problems, so additional application-specific information and logic are needed to translate the relatively tidy set of inputs seen in tables 6.1 to 6.6 into fully fleshed out input data for each tool.

Despite the required effort, many unique structural weight estimation programs have been created and described in literature, several of which were mentioned in section 2.2. The diversity of these programs suggests that it would be impossible to develop a weight estimation program that satisfies the needs of all possible vehicle concepts, stages of design, structural technologies, and experimental goals. Since it seems likely that there will be a continual need for ad hoc structural weight estimation programs, there is substantial opportunity for development time savings by providing the common functionality and interfaces needed by these programs in a software library.

This motivated the creation and continual development of the Rapid Airframe Design Environment (RADE) [133]. RADE, of which the author is a co-developer, is a software toolkit written in Python [75] for the purpose of making integrated aircraft structural weight estimation programs. RADE provides functions implementing domain-specific logic useful for such a program. RADE is also aware of some commonly used third-party software tools, and provides functions for assembling their inputs, invoking them, parsing their outputs, and transforming those outputs for downstream use. Ideally, a project that needs physics-based structural weight estimates can leverage RADE's existing capabilities to satisfy its typical requirements, and only have to develop new code to address its unique aspects.

The physics-based computational environment for the example use case was created by modifying an already available RADE-based CRM weight estimation program. This program uses a symmetric half-span finite element model of the structural wingbox and a

vortex lattice aerodynamic model that includes the wing and horizontal tail. A high-level flow chart of this computer program, which takes the specified parameters and choice variables as input and returns the structural weight as output, is shown in figure 6.3.

Outer Mold Line Geometry

The environment uses OpenVSP [134], a NASA-developed aircraft CAD tool, to handle the outer mold line geometry. OpenVSP allows aircraft geometry to be expressed through convenient high-level quantities such as aspect ratio, sweep, taper ratio, airfoil shape, etc., rather than in a more primitive form such as the coefficients of a NURBS surface. The baseline vehicle configuration, which is the geometry shown in figure 6.1, is stored in an OpenVSP file as separate geometric objects for the wings, horizontal and vertical tails, fuselage, and engine nacelles. Separating the parts of the vehicle helps later on when making vortex lattice surfaces for the wing and tail, when placing structural geometry in the wing, and when locating the engine's center of mass.

In each run of the environment, RADE directs OpenVSP via an API to change the baseline outer mold line so that it has the desired wing area and aspect ratio. As a heuristic to adjust the aspect ratio, only 5% of the normalized change in span comes from the section inboard of the yehudi break, while 95% comes from the outboard section. After the outer mold line is finished the geometry gets exported in a STEP formatted file, and then RADE loads the geometry from the file into memory for use in the next step.

Structural Geometry

The structural geometry consists of the main wingbox and a simple tail panel. The physics-based environment takes advantage of an available RADE function that creates a wingbox for conventional tube-and-wing aircraft. This function uses AFEM [135], a Python package for the flexible creation of aerospace structural geometry. AFEM in turn is built on top of geometric and topological objects defined by Open CASCADE Technology [80], a

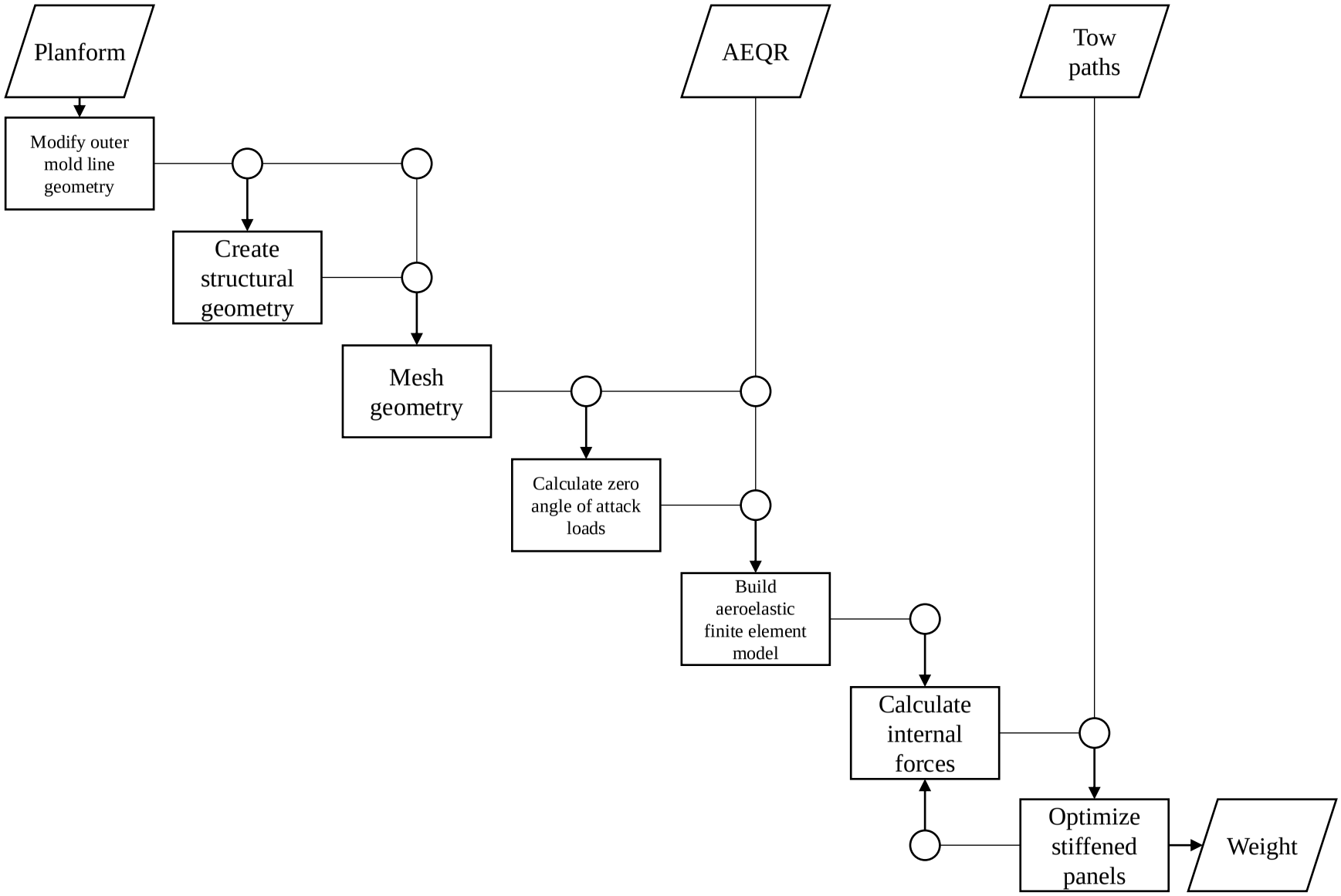

Figure 6.3: High-level overview of example use case's physics-based model

commercial CAD package. As mentioned in the formulation section, geometry creation can be brittle; an advantage of using this function to create the geometry is that it has already been stress tested on other configurations and should be all the more robust for it.

The wingbox-creation function is parameterized by the chordwise and spanwise extent of the wingbox and the spacing between ribs, allowing it to construct a variety of structural layouts. It is also able to kick the rear spar at the designated yehudi location. In this example use case the wingbox's structural layout variables are kept constant, but it would be easy to modify the sizing program to expose them if one wanted to also investigate the impact of changing the structural layout.

Using AFEM, the function creates surfaces representing the spars and ribs at the specified locations, and intersects them with the wing OML solid to bound them within the wing. It creates surfaces for the skins and fuses all the shapes together so that they have the correct topological connections; this also results in each skin panel and spar panel (a segment between two adjacent ribs) being represented by a distinct face. Finally, the portions of the ribs and skins in front of the front spar and behind the rear spar are discarded, leaving just the main wingbox. Panel stiffeners are not modeled geometrically, but rather through the element stiffness properties, and details like cutouts and access panels are not included at all.

Another RADE function is used to create a simple flat panel as the horizontal tail's structure. The purpose of this panel is to receive trim loads from the aerodynamic model's tail to balance the force on the wing, and not to accurately estimate the tail structure weight, so a simple panel is sufficient for this purpose.

Figures 6.4 and 6.5 show the wingbox structural geometry sitting inside the outer mold line geometry at the baseline and highest aspect ratio configurations, respectively. Figures 6.6 and 6.7 show closer views of the wingbox geometry.

After all these shapes are created, they are organized hierarchically by region and part group (e.g., a face on the rear spar could be accessed, with zero-based indexing,

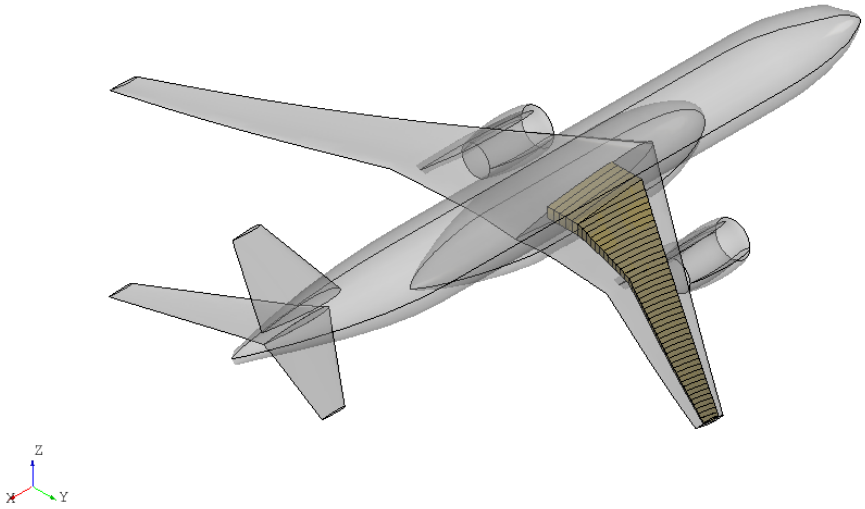

Figure 6.4: Outer mold line and structural geometry for baseline CRM

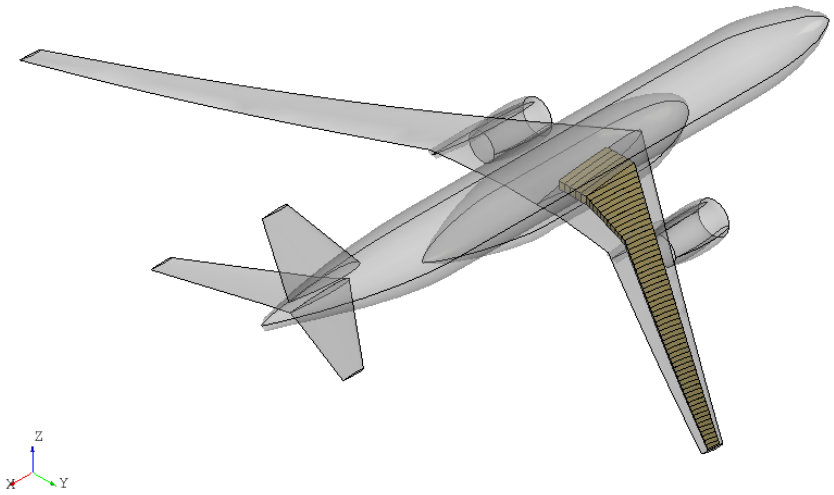

Figure 6.5: Outer mold line and structural geometry for high aspect ratio CRM

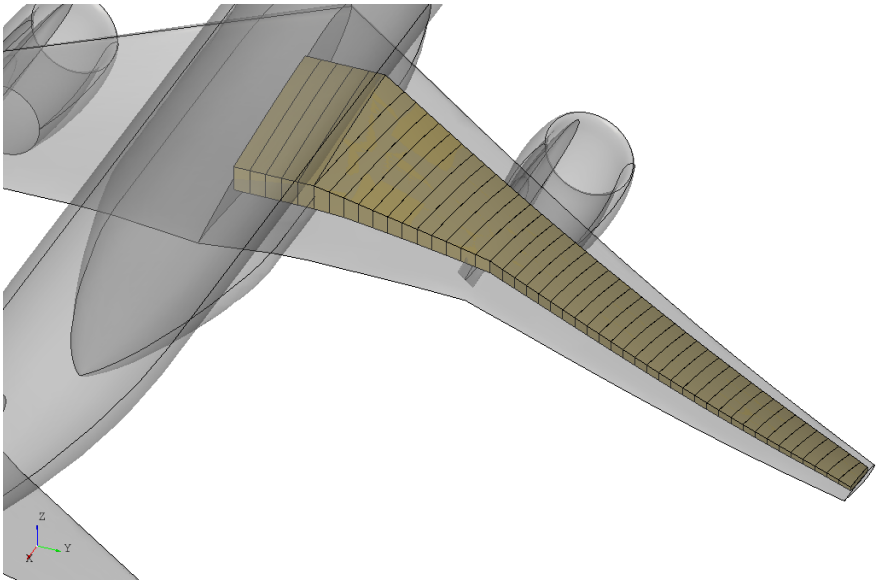

Figure 6.6: Wingbox geometry for baseline CRM

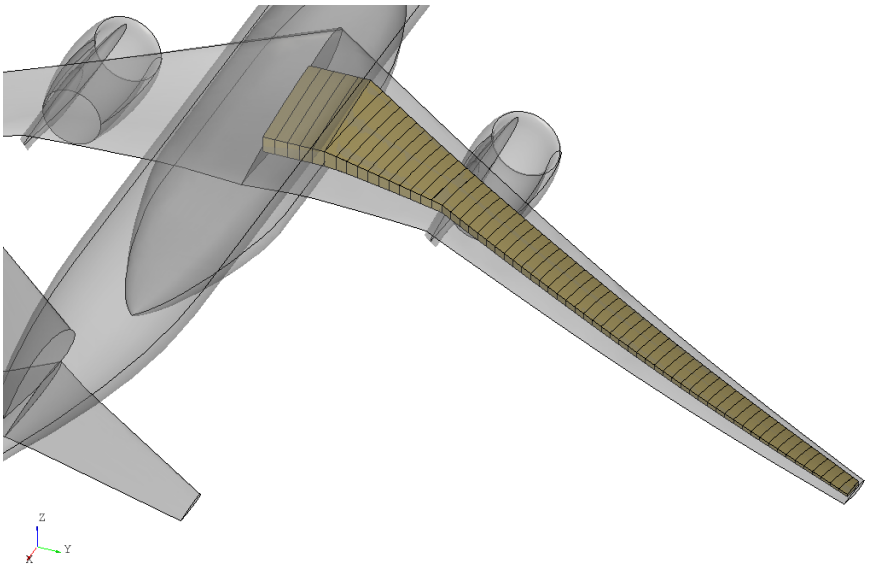

Figure 6.7: Wingbox geometry for high aspect ratio CRM

via “wing”-“spar”-1-14). One of the author’s contributions to RADE was a set of classes following the composite design pattern from software engineering to provide uniform access to the underlying shapes, regardless of whether a component of the hierarchy organizes its subcomponents by sequential indices, by named keys, or is itself a leaf node with no subcomponents. This capability to uniformly access all of the contained shapes helps with downstream processes like assigning stiffness properties to a group of related mesh elements.

Lastly, another RADE function is used to calculate the volume and centroid location of each rib bay. This information will be used to add the mass of fuel within each bay to the structural model.

Mesh

A couple of meshes are needed for the physics-based model. The structural finite element mesh can be created with RADE, which in turn uses SMESH [81] and Netgen [136]. Another one of the author’s contributions to RADE was a set of convenience wrappers for SMESH’s mesh, element, and node data structures. A length of seven inches was selected for the nominal mesh spacing. Although in most cases the mesh algorithm is able to produce a good quadrilateral mesh, for triangular faces and for quadrilateral faces with excessive angles it tends to produce mesh elements with poor aspect ratios and interior angles, so a triangular mesh is used instead. This strategy allows for a superior quadrilateral mesh throughout most of the structure, but with the robustness to still provide acceptable meshes in difficult regions. The resulting elements are assigned orientation angles that are aligned with the rear spar for the skins, and vertically for the spars and ribs. These orientations do not rotate or otherwise geometrically affect the elements, but rather are used later to ensure the homogenized stiffness properties reflect the correct stiffener alignment.

The mesh for the vortex lattice aerodynamic model is created by RADE directly. AFEM is able to extract the planform surface from the wing and tail outer mold lines and RADE discretizes that surface to form the vortex lattice mesh. The mesh length is set so that there

are ten chordwise elements along the wing.

Zero Angle of Attack Loads

The physics-based environment accounts for the camber of the wing's airfoil sections by calculating a set of zero angle of attack loads with AVL [137], a vortex lattice software. Another one of the author's contributions to RADE was a module for preparing AVL input files, calling the executable, and parsing the output files. The zero angle of attack loads can then be applied to the aeroelastic finite element model. Although the camber could also be directly accounted for in the aeroelastic finite element model, it is convenient to do so in this fashion because AVL allows for the camber to be input using the airfoil cross sections. This eliminates the need to write additional code to calculates the proper pitch angle for each doublet lattice element from the outer mold line geometry.

Build Finite Element Model

RADE then assembles all of the previously generated information about the meshes, loads, and the case's AEQR setting into a bulk data file to use as an input to Nastran [76], a commercial finite element solver. RADE is very useful at this stage for transforming all the data into the needed format. For instance, the structural mesh was made with SMESH and Netgen, and is represented using SMESH's in-memory format. However, Nastran needs the mesh to be input in its own text format and combined with information that SMESH knows nothing about, like elastic properties, mass properties, and boundary conditions. RADE provides functions to collect all this information from the different sources and aggregate it in the proper form.

At this stage some additional grid points are created to help with boundary conditions and connections. These points are created by fitting a line through the centroids of all the ribs and taking the intersections between this line and each rib. These grid points do not correspond to a literal part of the structure, but provide an abstraction of each rib's average

displacement and facilitate the application of external loads to each rib's entire cross section. This is accomplished by connecting these grid points to the edges of their associated rib using rigid interpolation constraint elements (RBE3 elements). These elements force the added grid points to move with the average displacement of the rib-edge grid points, and reciprocally distribute forces on the added grid point to the rib-edge grid points. Using this system of rib reference points, the zero angle of attack loads on each rib can be specified by a simple force/moment pair that is automatically distributed over the entire cross section.

Masses representing the wing fuel and the engine are added to the model. The volume of each rib compartment is multiplied by a fuel density of 6.84 $^{lb}/_{gal}$ to calculate the mass of fuel it contains, and concentrated mass elements are placed at the rib-bay centroids and connected to the surrounding structure with rigid interpolation elements. The engine mass (including the nacelle and pylon) is placed at the engine's geometric center and is connected to the nearest two ribs with more rigid interpolation elements.

A final grid point is added at an assumed aerodynamic center location, such that the vehicle's static margin is 0.1. A rigid element (RBE1) joins the root rib of the wing and the root edge of the tail to this point. This serves to join the wing to the tail and is where a load representing the rest of the aircraft's weight is applied. This point is constrained in the streamwise and lateral directions but unconstrained in the vertical direction. The vertical direction will be constrained by an acceleration requirement that will be resolved by trimming the model's aerodynamic forces.

Each finite element is assigned initial mechanical properties. These properties will soon be updated to reflect the local panel designs, but initial values are needed to perform the first finite element analysis to get the first set of internal forces. The initial properties are representative of aluminum-like stiffness and density. The elements are grouped by topological face so that the each skin and spar section within a rib bay has its own property definitions. Since each element within a group has the same properties, and because the properties are representative of the design, this discretizes the design by topological face.

Therefore the panel thickness, stiffener height, etc. are constant within each face.

A vortex lattice model is used to calculate the lift force on the wing and horizontal tail. RADE is able to extract the planform surface from the wing and tail outer mold lines and discretize it to form the vortex lattice surfaces. To capture the aeroelastic interaction between wing deflection and aerodynamic forces, these surfaces are connected to the structural model by a beam spline at the aforementioned rib reference grid points. The spline transfers the average motion of the rib cross sections to the vortex lattice model's panels and transfers load from the vortex lattice model's panels to the rib cross sections. The AEQR parameter directly controls this interaction, and is defined as part of Nastran's trim inputs. If AEQR is set to one then the trim procedure occurs normally and the aeroelastic interaction is modeled. If AEQR is set to zero then the forces are still transferred from the aerodynamic model to the structural model and the aerodynamic model is still trimmed, but the structural displacements are not sent to the aerodynamic model. When AEQR is zero the aerodynamic forces only depend on the trim angle of attack and not on the deflection of the structure. Fractional values of AEQR transfer a fractional amount of structural displacement to the aerodynamic model.

The model will be analyzed at the pull-up and push-down load cases from table 6.7. The desired load factors are obtained by trimming the angle of attack and a net zero pitching moment is attained by trimming the horizontal tail pitch angle.

Calculate Internal Forces

At this point there is a large amount of data in memory describing the aeroelastic model. In order to analyze it RADE translates the data to a text formatted bulk data file with the help of the nastran_utils package [138], and also adds the instruction to perform static aeroelastic analysis. The Nastran executable is then called with the bulk data file as input to carry out the analysis, resulting in the structure's internal loads.

This step is repeated within an execution of the weight estimation environment because

it must be iterated with the stiffened panel optimization step that follows. Each time, after the panel designs are updated, the model's mechanical properties also updated, the input file is rewritten, and the analysis is redone to obtain a new set of internal loads.

Optimize Stiffened Panels

The actual structural sizing is carried out by iterating between Nastran and HyperSizer [132] in a process referred to as the HyperFEA loop. This is an implementation of a fully stressed design type approach to structural optimization. At a high level, Nastran solves the aeroelastic system of equations to calculate the internal loads in each finite element, and then HyperSizer selects stiffened panel designs to minimize the weight of the components without failing due to those internal loads. After optimizing the panel designs, the finite element properties are updated to reflect those designs. This approach to resolving the coupling between the internal forces and panel designs is an example of fixed point iteration.

The author created the RADE module responsible for providing an interface to HyperSizer and directing its use for common workflows. The following paragraphs provide details about this process.

The HyperSizer application is primarily intended to be used via a gui, but to enable automation it also provides a COM interface. Some setup is required before HyperSizer is able to design the structure. After starting HyperSizer RADE directs the creation of a new project, and then imports both the finite element model and its results from the bulk data file and Nastran's binary output file (.op2), respectively. During the import HyperSizer automatically assigns groups of elements to distinct structural components based on their property group. It also automatically detects the length, width, and curvature of each component based on the geometry of its elements, which can then be used buckling failure calculations.

Next the panel concept for each component is assigned, which in this case is the blade-stiffened panel concept from figure 6.2. It has dimensions for skin thickness, panel height,

web thickness, and the distance between stiffeners. The skin stiffener spacing was fixed at six inches because the same skin stiffeners extend through all the rib bays and should have consistent spacing. For the spars and ribs, which have vertically aligned stiffeners, the stiffener spacings are free to be optimized.

After selecting the panel concept, the materials for the skin and stiffener are assigned. Each component uses the composite laminates described in tables 6.4 to 6.6. HyperSizer provides two different representations of laminates: effective laminates and discrete laminates. The effective laminate provides a simpler representation using only thickness and the fractions of 0°, ±45°, and 90° plies. Effective laminates were used in all the structure except for in the skins. Effective laminates were not suitable for the wing skins because their restricted ply orientations are unable to model a tow-steered structure. Discrete laminates, which take the straightforward approach of stacking an arbitrary sequence of plies together into a laminate, were used instead. Tow steering was implemented by assigning discrete laminates with the specified orientation to each of the wing's skin panel sections. Laminates were composed by repeating a core sublaminate. Different numbers of repeats provided different thickness options, and rotation of the core sublaminates provided the desired steering angles. The core sublaminate had a stacking sequence of $[\pm 45°, 0°_4, 90°, 0°]_S$ to achieve the ply fractions in table 6.4.

After the panel optimization problems have been set up and the internal forces have been imported, it is finally possible to find the panel designs. HyperSizer is intended to be used to optimize structural material selection and other fundamentally discrete variables like the number of plies in a laminate, and as a result it uses a brute force optimization algorithm that, while crude, covers all its use cases. This algorithm works by first creating candidate designs based on the user-selected options for panel concept, structural material, and a grid of values for each dimension. Next, these candidates are sorted from lightest to heaviest. Finally, the candidates are evaluated one at a time to determine if they have sufficient margins of safety for all of the failure analyses. The first candidate that does not

fail is chosen because it is the lightest-weight feasible candidate. HyperSizer then calculates the homogenized stiffness [139] for all of the new panel designs and exports this information in a new bulk data file that can be used by Nastran.

Although the brute force algorithm is robust, it is inefficient. If the panel dimensions are coarsely discretized then the selected design is likely to be significantly heavier than the truly optimal design. If the panel dimensions are finely discretized then there will be an exponentially large number of candidate designs and it will take a long time to find the solution. To help with this issue, the author contributed a grid search algorithm to RADE's HyperSizer module. This algorithm iteratively directs HyperSizer to solve the problem on a coarse discretization and then updates the discretization based on the solution. If a variable is set to the upper or lower bound of the grid, then the grid is shifted in that direction. If a variable is set to the middle of the grid, then the range is shrunk. This grid search greatly reduces the time it takes for HyperSizer to find a precise optimal panel design.

In order to perform the failure analyses, the internal loads from the finite element analysis must be mapped to the panel objects (the clear span and the stiffener web in this case). HyperSizer aggregates the element loads into component loads by taking the average plus two standard deviations of the load in each component's elements [140]. Separate component buckling loads are calculated in the same way but with compressive loads only. After aggregating the component loads they are split into object loads based on the stiffness of the individual panel objects. These loads are then increased by a standard ultimate factor of safety of 1.5.

HyperSizer can then evaluate the failure criteria based on these object loads. HyperSizer provides built-in support for many common analytical failure theories. The weight estimation environment uses: buckling [141] of entire panels [142] and local panel sections [143], crippling and crippling/buckling interaction [144], and Tsai-Wu composite strength [145].

Output Weight

After the panel designs and internal forces have converged the structural sizing is completed. The final step in the physics-based weight estimation is to report the final weight. This comes from the panel cross section dimensions, the panel areas, and the structural material density. Conveniently, HyperSizer already provides these calculations by part, by assembly, and over the entire structure. In addition to the total weight the program records the group weights for the upper skin, lower skin, spars, and ribs.

6.3 Initial Stage of Experiment

After preparing the physics-based weight estimation environment/objective function, the next step in the methodology is to run the initial stage of data collection. This example use case combined a 170-case maximin Latin hypercube over all the input variables with a 20-case maximin Latin hypercube over just the planform variables. For this secondary set of cases AEQR was fixed at 1 and all the tow orientation coefficients were fixed at 0°. The data from the second Latin hypercube will give the surrogate model a higher level of accuracy for designs with conventional straight span-aligned tows, which will be useful for comparing with the best tow-steered designs.

There were some issues with the automated geometry modification and generation that resulted in 4 out of the planned 190 cases failing. Since these failures happen early on in the physics-based model process the lost time is relatively insignificant. Although these cases could be made up by slightly perturbing the planform variables and trying again, this was deemed an acceptable number of failed cases. Since the rest of the data collection is adaptive, if the holes from the failed cases are in important regions of the design space those regions are sure to be revisited.

6.4 Validate Surrogate Model Form

The surrogate model form chosen for the example use case was a Gaussian process with a quadratic mean function and a Matérn ⅗ covariance function. The mean function has intercept, linear, and pure quadratic terms, but no interaction terms. This model form expresses the response as the sum of gradual global curvature and a rough nonlinear function. The roughness of the Matérn ⅗ covariance gives the surrogate more flexibility to model nonsmooth behavior without explaining it as noise, compared with the other stationary covariances. The hyperparameters were fit using REML.

The results from experiment 3 indicate that overfitting should not be an issue for this model form, as it uses no factor analysis directions and REML is used to estimate the hyperparameters. The initial leave-one-out cross-validation actual vs. predicted plot is shown in figure 6.8 and reveals that the initial surrogate model is capturing trends in the data but that there is still a need to improve its accuracy. The normal quantile plot in figure 6.9 indicates that the surrogate has a good understanding of its own accuracy since the standardized residuals fit a standard normal distribution very closely.

6.5 Sequential Stages of Experiment

An additional 170 cases were evaluated and added to the data set with Bayesian parameterized optimization. The Monte Carlo approximation of the integrated knowledge gradient used 1,000 samples with 30 unique parameter settings. As recommended, antithetic variates were used for the standard normal variable and Latin hypercube sampling was used for the planform samples. AEQR also needed to be sampled because it is part of the parameter set, but because of its special role it was only sampled at the values of interest 0 and 1. This corresponds to using a discrete uniform rather than a continuous uniform probability distribution for this variable. $\mathbb{IKG}$ was maximized by multistart gradient-based optimization with 34 random starting locations.

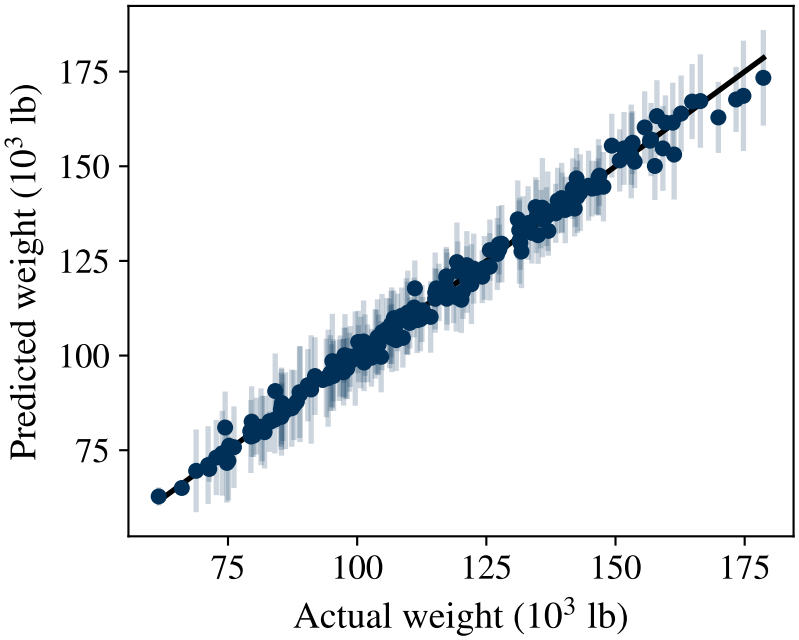

Figure 6.8: Example use case initial actual vs. predicted plot. The error bars indicate three standard deviations above and below the prediction mean.

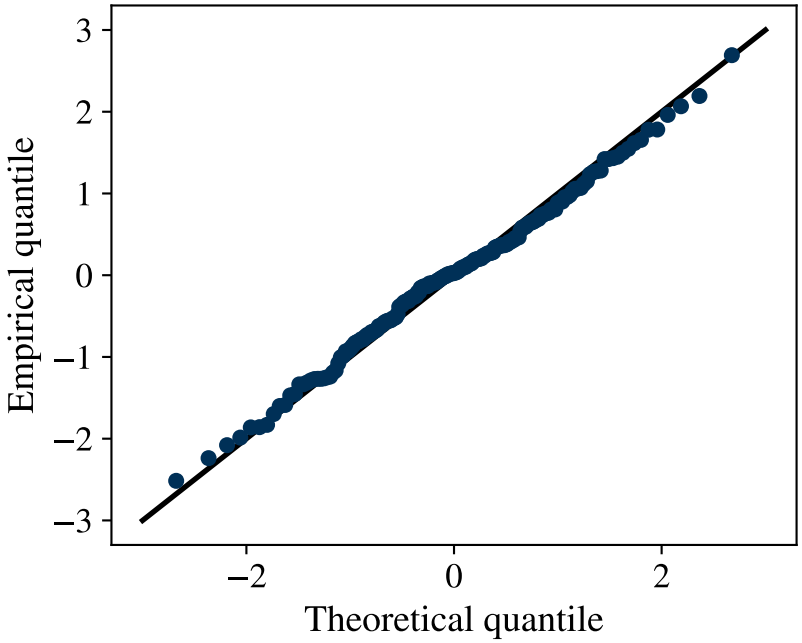

Figure 6.9: Example use case initial normal quantiles plot

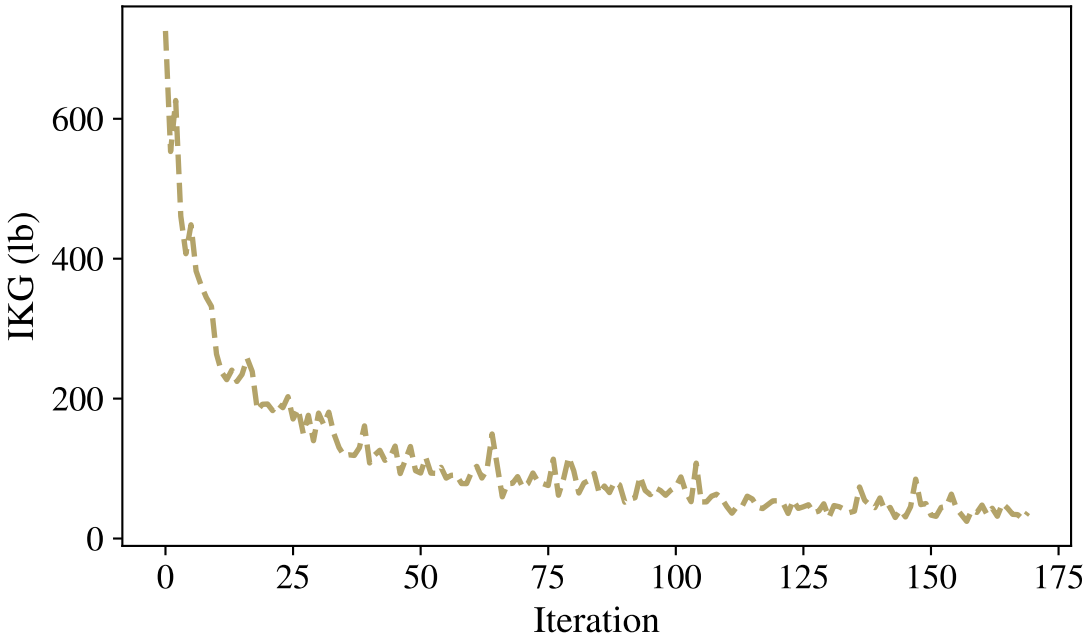

Figure 6.10: Evolution of IKG in the example use case

There were six cases that failed for geometry-related reasons in this phase of data collection. These failures were easily dealt with by retrying at the next-best local maximum found during the multistart $\mathbb{IKG}$ optimization. Like in the first stage, these failed cases did not cause much time loss because the failures occurred early in the physics-based procedure.

Figure 6.10 shows how the maximal $\mathbb{IKG}$ decreased over the course of the sequential data collection stages. It started around 730 lb and fell to around 40 lb at the end. This is higher than the final $\mathbb{IKG}$ values seen in experiment 3, which suggests that improvement is still possible, but it is also clear that the rate of potential improvement has slowed significantly.

The leave-one-out actual vs. predicted plot after all the data has been added is shown in figure 6.11. Line in experiment 3 there are indications that Bayesian parameterized optimization has worked correctly: the newly added data has smaller error bars associated with it and it also tends to sample lower weight designs. Figure 6.12 illustrates this second point more clearly by showing a histogram of the samples' structural weight.

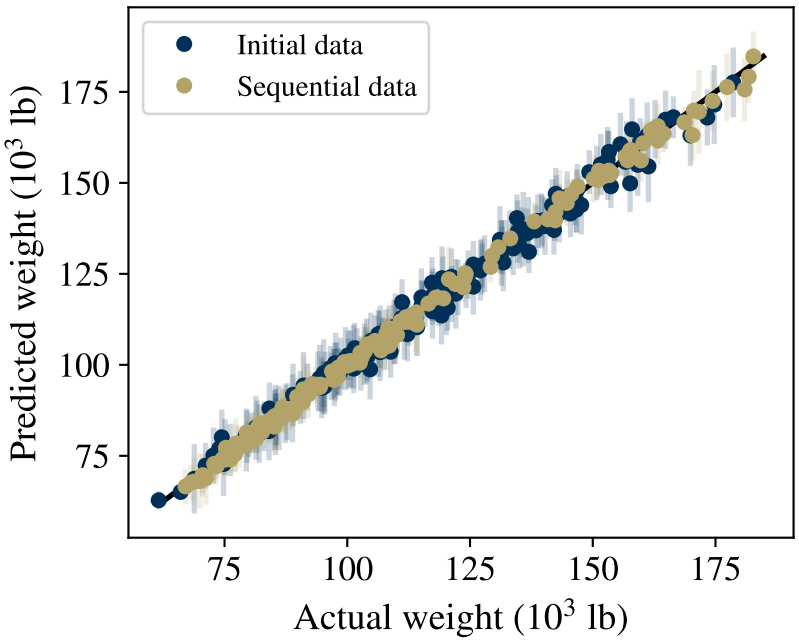

Figure 6.11: Example use case final actual vs. predicted plot. The error bars indicate three standard deviations above and below the prediction mean.

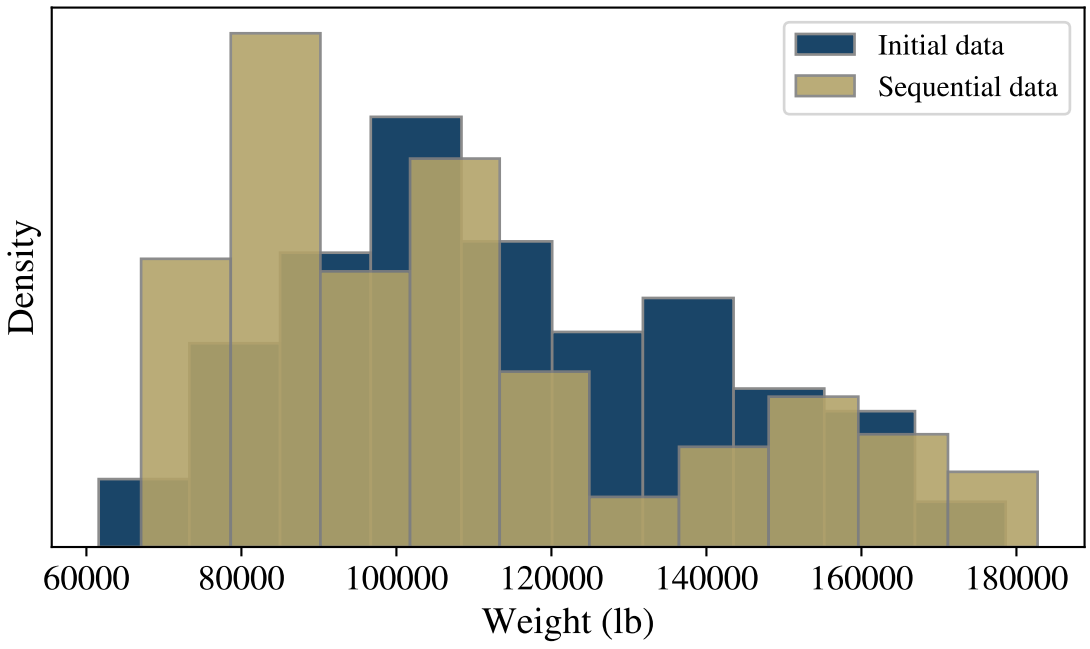

Figure 6.12: Sample density of structural weight in example use case

6.6 Validate Surrogate Value Function

To validate the accuracy of the final surrogate model at its predicted minima, it was optimized at twenty maximin Latin hypercube points that covered the range of planform variables and had an AEQR of one. These predicted minima were then evaluated on the physics-based model. Even though the methodology does not call for it, these same points were also optimized on the initial surrogate model and evaluated on the physics-based model. This provides a comparison of both the predicted minima and the surrogate accuracy before and after the adaptive data collection step. Note that error being measured is different from the error in the surrogate value function. Although that error would be more relevant to the problem, there is no way to directly measure the true value function to make the comparison.

Figure 6.13 shows the error observed in these validation cases relative to the actual weight. Both the initial and final surrogate model's minima are biased downward, but while the initial surrogate model's errors are around eight percent, the final surrogate model's errors have been reduced to around one percent. More detail can be seen in figure 6.14, which shows an actual vs. predicted plot that includes both sets of predicted minima. The thin gray lines connect points that were optimized at the same parameter locations but on different surrogate models. All of these lines slope diagonally up and to the left, which indicates that the initial surrogate model always predicted a lower minimum than the final surrogate model, and was always less accurate. It is also apparent how much the accuracy has improved.

After collecting this validation data, the accuracy of the final surrogate model can be approximated to be around one percent. This new validation data can also be added to the surrogate model to further improve its accuracy, so the analysis of the results in the next section should be at least this accurate.

The flatness of the acquisition function suggests that removing the remaining error would be difficult, and this level of error may in fact be at the level of the irreducible noise in the

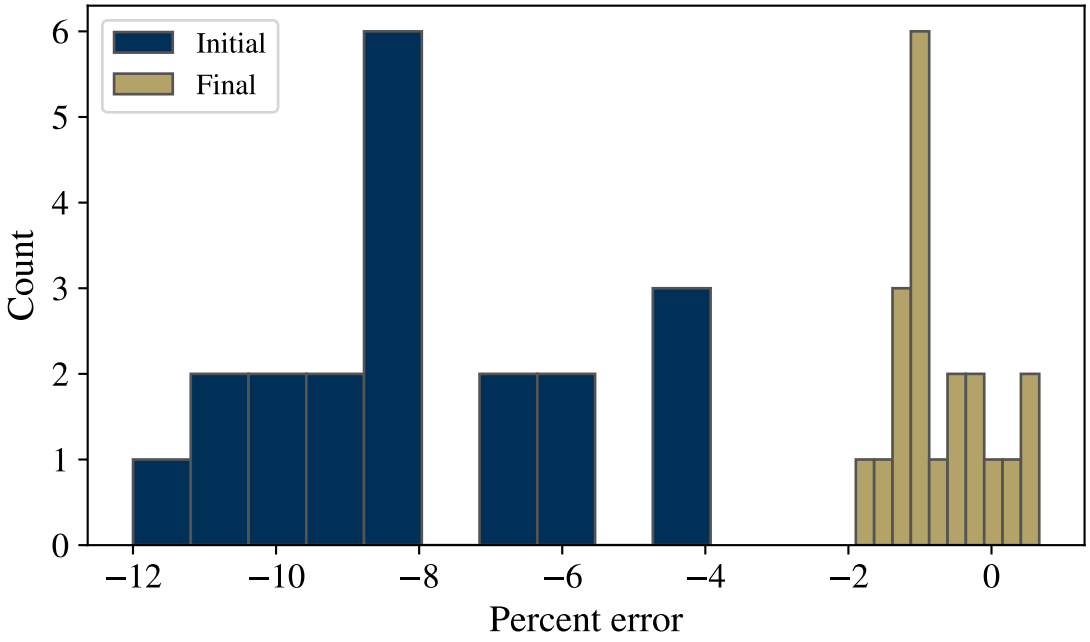

Figure 6.13: Relative error of surrogate model optima

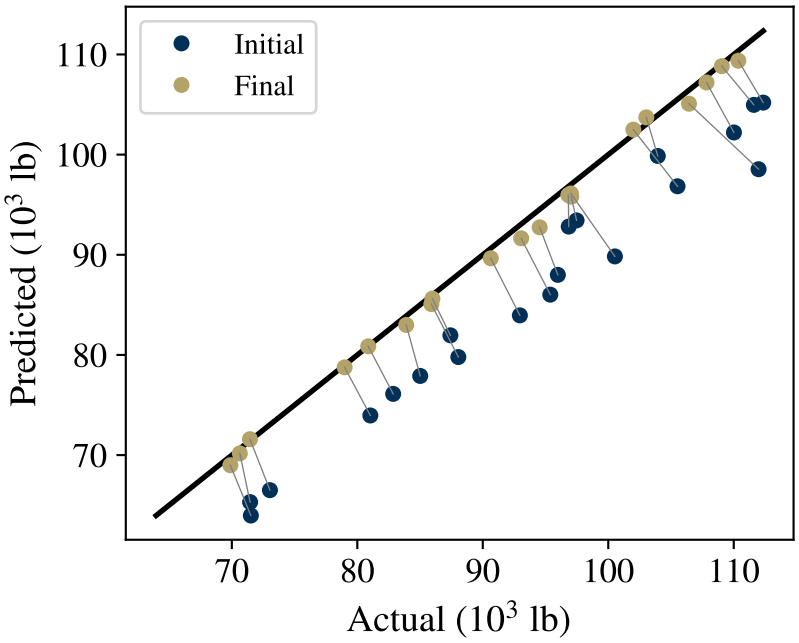

Figure 6.14: Actual vs. predicted of surrogate model optima. The thin gray lines connect points that represent the same location in parameter space but are different based on which surrogate model was used.

physics-based model. Noisy behavior of the physics-based model is a result of: termination criteria for the local structural optimization and the fully stressed design iteration, discrete changes in the number of ribs as the planform varies, and discrete changes in the finite element mesh as the planform varies. Measuring the magnitude of this noisy behavior was not done as part of this study, but it would be possible. The amount of noise due to the termination criteria could be estimated by running additional cases that perform a few more iterations of either the local structural sizing or the fully stressed sizing/loads iteration, depending on which is being measured. The noise caused by discrete changes in the number of ribs or the mesh could be estimated by first identifying special settings of the planform parameters for which a slight perturbation of that setting triggers the discrete change. The noise level could then be estimated by running the cases on each side of the discontinuity and comparing their results.

The accuracy of the final surrogate model over the validation cases complements the results from chapter 4 by giving further support to hypothesis 2. Unlike in experiment 3, in this use case Bayesian parameterized optimization was applied to a more realistic structural sizing simulation with a representative outer mold line geometry, stiffened panels, and stability-based failures. Despite these differences, the method still produced an accurate surrogate model.

6.7 Results and Discussion

Hyperparameters

The hyperparameters of the final surrogate model (including the validation cases) are shown in figures 6.15 through 6.17. There are strong linear and quadratic effects from AEQR; together they correspond to a typical difference of 34,000 lb in structural weight from choosing whether or not to include aeroelastic interactions in the physics-based model. Without accounting for these interactions the weight will be overestimated by ~30%. There is also a strong linear effect for increased weight with aspect ratio.

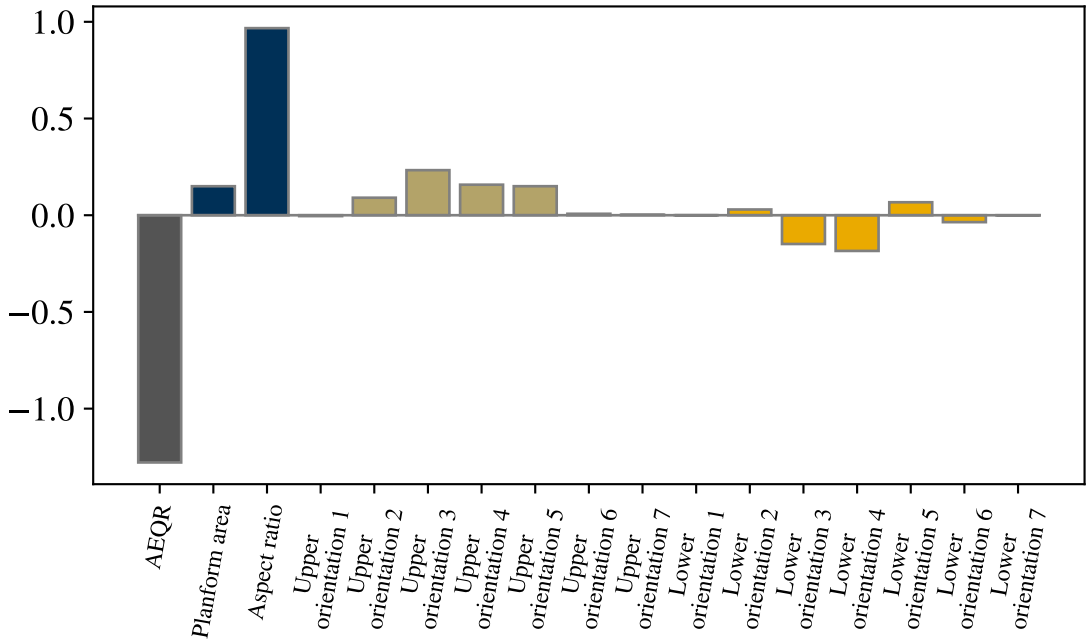

Figure 6.15: Centered and scaled linear coefficients of mean function

The linear effects for the tow orientation coefficients are all small and the quadratic effects are positive and largest in the middle of the span. This suggests that the best tow paths are likely pretty close to the zero direction (pointing along the rear spar). Note that these scaled values all depend on the corresponding variable's design range, so the inboard tow orientations' effects appear weaker due to their smaller design range.

It is also worth pointing out that the root tow orientation coefficients seem very week, but that this is probably due to modeling choices more so than physics. For convenience, the physics-based model defines the tow orientation splines as extending from the centerline to the wingtip, but the tow steering is not actually applied to the center body wingbox. Therefore the root coefficient affects a smaller portion of structure than the others and its role is diminished.

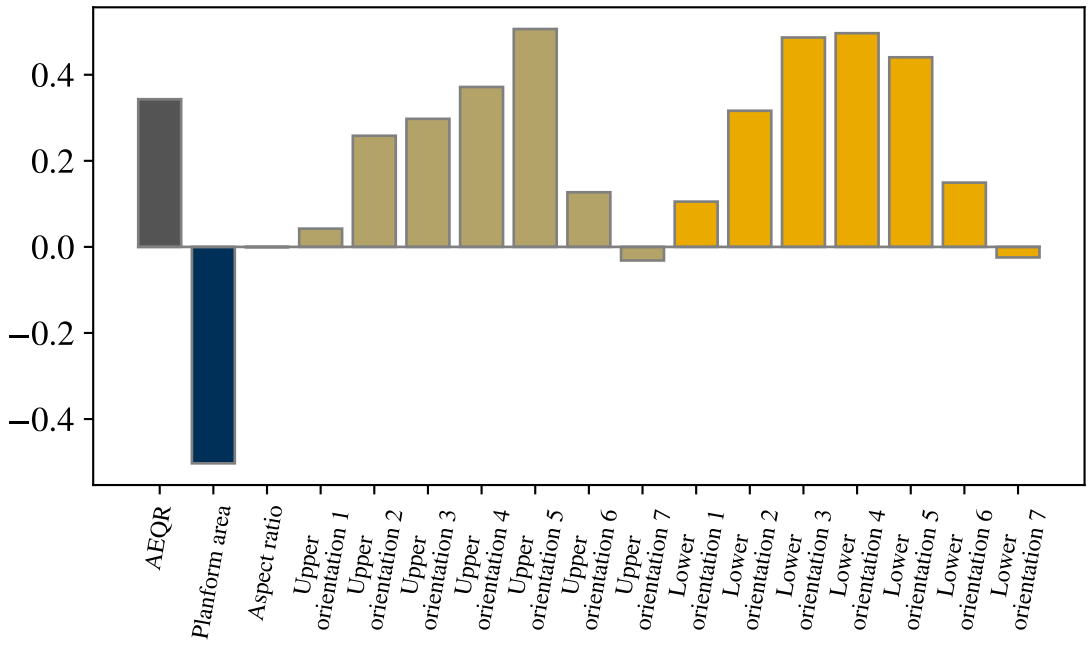

Figure 6.16: Scaled quadratic coefficients of mean function

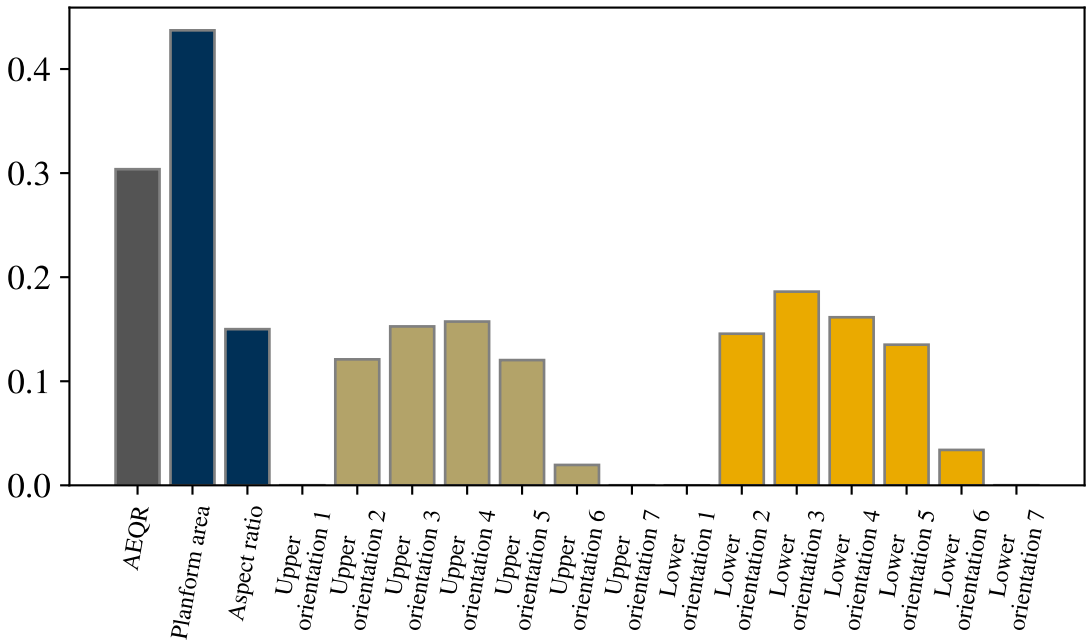

Figure 6.17: Scaled covariance sensitivity hyperparameters

Technology Benefit

In order to determine the technology benefit provided by tow-steered composites the final surrogate model was optimized throughout the planform space. First, an optimal straight-tow design was obtained by rotating each skin's fiber orientations as a single unit, which was achieved by using a single design variable to control all the spline coefficients on each skin. The resulting design has straight tow paths that are not necessarily aligned with the rear spar. Next, the best tow-steered design was found by optimizing with all the tow path degrees of freedom. To give a good chance of finding the global optimum even if there are multiple local minima, the optimization was performed with the basinhopping algorithm from Scipy.

Contours of the best weights for straight-tow and tow-steered designs are displayed in figure 6.18. The behavior of the two are very similar, with nearly linear increases in structural weight with area and aspect ratio. This is interesting because the sensitivity hyperparameters indicate that there is nonlinearity associated with these variables in the objective function, but that does not translate to nonlinearity in the value function. This means that it can be relatively easy to extrapolate what the optimal structural weight will be at a new planform, but much harder to determine the structural design associated with that optimal weight. For the design ranges studied aspect ratio has a stronger effect.

The difference in optimal weight between straight-tow and tow-steered designs is subtle. In order to see the difference more clearly, figure 6.19 shows contours of the percent difference of optimal tow-steered weight relative to optimal straight-tow weight. This quantity is the weight-reduction factor sought by the research objective, and could be added to a system-level vehicle model in order to account for the use of tow-steered composite technology. The improvement ranges from only around a 0.6% weight reduction to around a 3.5% weight reduction. The response is relatively flat and changes gradually across most of the planform space. However, the benefit from using tow-steered composites is noticeably greater at low area and high aspect ratio. This corresponds to vehicle designs with higher wing loadings and larger root bending moments, which indicates that tow steering tends

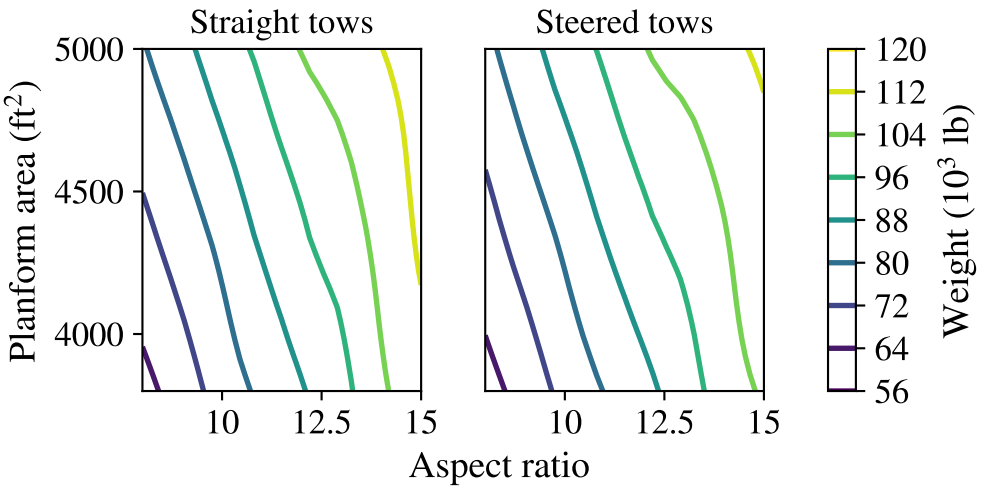

Figure 6.18: Contours of optimal structural weight for straight and tow-steered designs

to be more beneficial for vehicle designs that put more challenging loads on their wings. Although the variation in the technology benefit is small, its behavior is distinctly nonlinear. Low-order regression models, like quadratic response surfaces, would be unable to represent it. This justifies the use of a more flexible surrogate model form such as the Gaussian process employed here.

The size of the weight reduction benefit is smaller than other values reported in literature. For instance, Brooks, Martins, and Kennedy [41] also examined the effects of tow steering on a high aspect ratio CRM configuration and found that it provided a nine percent reduction in structural weight relative to a straight-tow design. They mostly attributed this benefit to localized effects, like improved resistance to loads and structural failure, rather than to overall wing effects like passive aeroelastic load alleviation. This assessment is similar to those made by other researchers [146]. Unfortunately, since the physics-based model used here does not provide derivatives, it is not possible to directly compare the results from the thesis methodology with those from a high-dimensional gradient-based optimization approach, which would help to diagnose why the benefit is different.

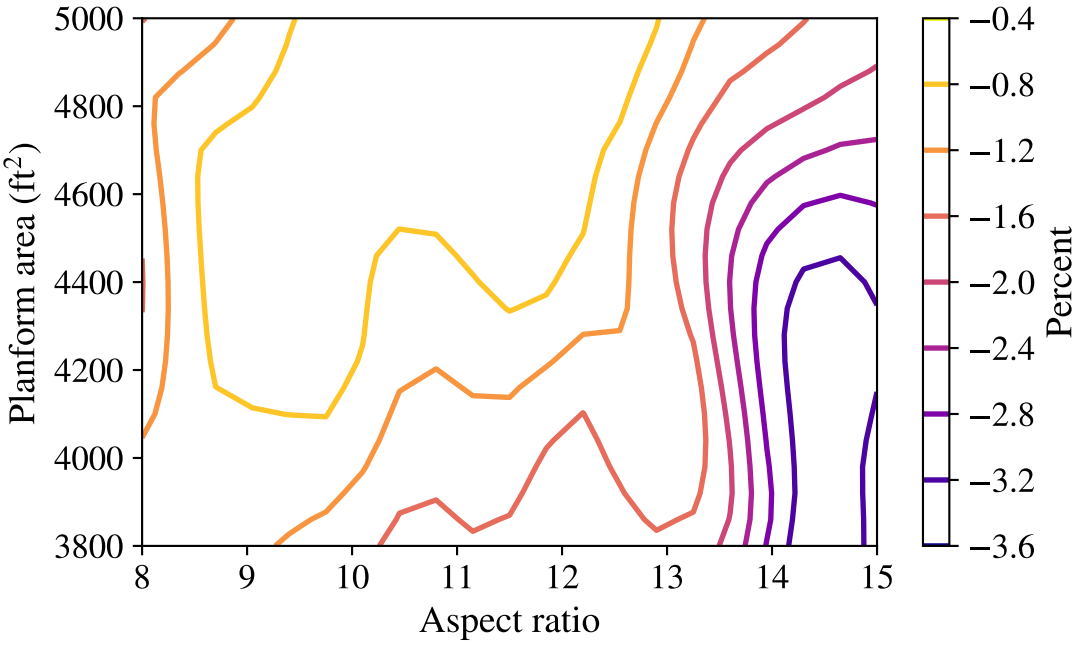

Figure 6.19: Percent difference of optimal tow-steered weight relative to optimal straight-tow weight

A potential explanation is that rapid changes in the tow orientations are necessary to gain a large portion of the maximum technology benefit, and the thesis methodology uses too coarse of a parameterization to obtain that portion. If this is the cause then that means that the coarse tow path parameterization used in the methodology is only able to capture about a third of the potential benefit, and additional measures, such as applying a supplementary adjustment factor, are needed to estimate the total technology benefit. There are other important details that could be partially responsible for the difference between the results found here and in literature. First, tow steering was not applied to the center fuselage carry-through section of the wingbox, so potential weight reduction benefits for a significant portion of the overall structure were left on the table. Second, the approach in [41] allowed free form deformation of the wing's outer mold line, which provides substantially more design freedom.

A cross section through the technology benefit contour is taken at a planform area of

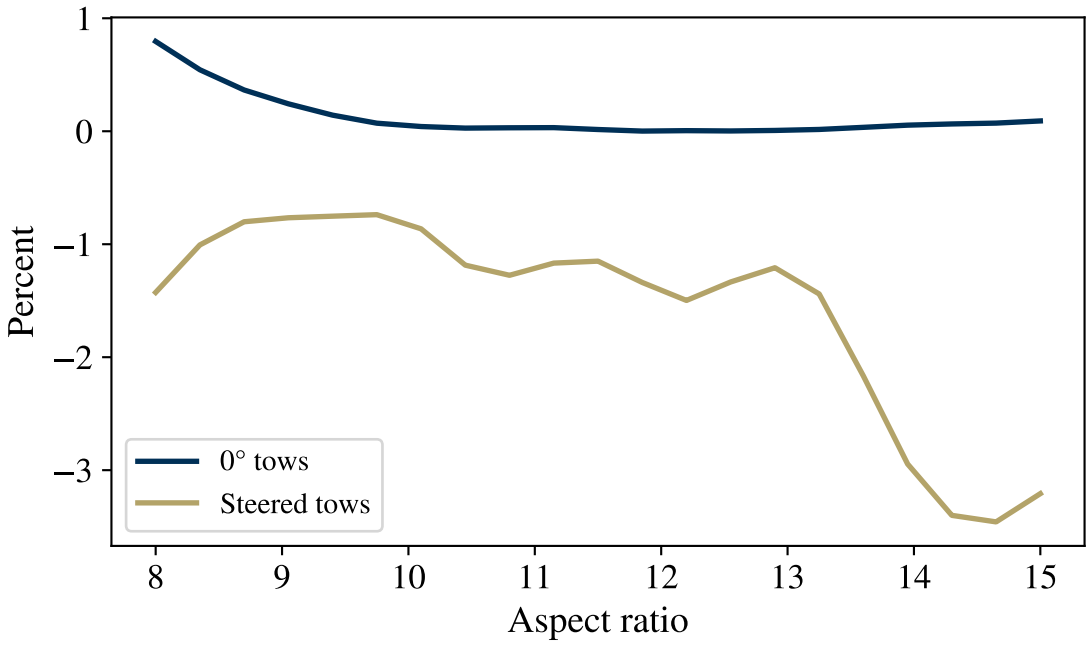

Figure 6.20: Percent difference from straight-tow design at planform area of 4,160 ft^2

4,160 ft^2 in order to form figure 6.20. This gives a better view of the relative flatness and more sudden decrease. There are a couple small oscillations in the flat region, though they are within the accuracy limit of the surrogate so they might be spurious. These oscillations are also visible in the previous image as the cusps appearing between aspect ratios of 10 and 13. If they are not spurious, they could indicate that there are multiple local minima, and that the global minimum suddenly switches from one basin to another at those parameters.

This figure also shows the difference between the optimal straight-tow design and the zero-degree design with all tows pointing along the rear spar. Although the zero-degree design is a slightly heavier at low aspect ratios, the two are practically identical when the aspect ratio is greater than ten.

A comparison of the tow orientations themselves is presented in figure 6.21, which shows the upper and lower orientation angles as a function of normalized span location for a sweep of aspect ratios. The most noticeable trend in these plots, the extreme angles

at the wingtip, are visually overemphasized. As the surrogate's hyperparameters indicate, the wingtip tow angles do not have a strong effect on the structural weight. This pattern is merely due to the optimizer pushing variables with insignificant but nonzero effects to their bounds, quite banal.

A trend that is interesting and important is the difference in behavior between the two highest aspect ratio designs and the others. These higher aspect ratios correspond to the region of planform space where the improvement from tow steering is much greater. On the lower skin, the high aspect ratio designs have negative orientations near the root and positive orientations in the outboard section, while the low aspect ratio designs stay much closer to zero. There are similar trends on the upper skin, but for the outboard section the designs form more of a gradient rather than a high aspect ratio group and a low aspect ratio group.

Global vs. Local Optimization

To further study the features from figures 6.19 and 6.20 that suggest multiple local minima, contours of the optimal setting for one of the most important choice variables, the third lower skin tow orientation coefficient, are shown in figure 6.22. Like in figures 6.19 and 6.20, there are cusps in the region of lower planform areas and aspect ratios between 10 and 13, which might indicate multiple local minima. Surprisingly, there is a discontinuous jump in the optimal setting at the lowest aspect ratios and an area of around 4,750 ft^2, which is a location that had no interesting behavior in the other figures. However, this is clear evidence of different optimal basin on either side of the jump.

To examine the consequences of using local rather than global optimization, the surrogate was optimized using a gradient-based algorithm started from the zero degree straight-tow design. A comparison of these designs with the globally optimized designs is in figure 6.23. For almost the entire planform space there is no difference, but there is a noticeable region of worse performance in the same place that lower orientation coefficient 3 has its discontinuity. This provides another indication that there are multiple local minima at this location, but

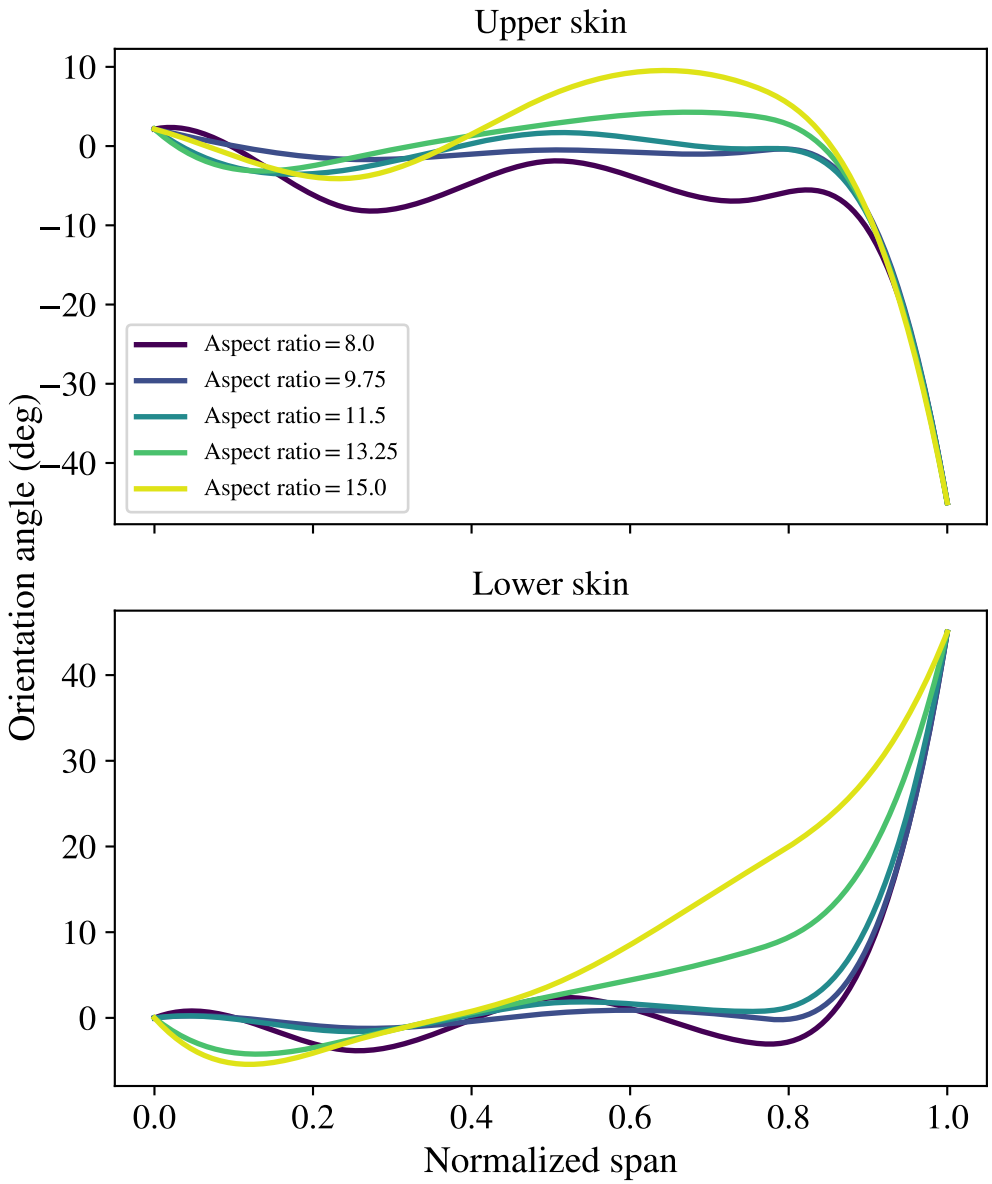

Figure 6.21: Optimal tow orientation angles for a sweep of aspect ratios at a planform area of 4,160 ft^2

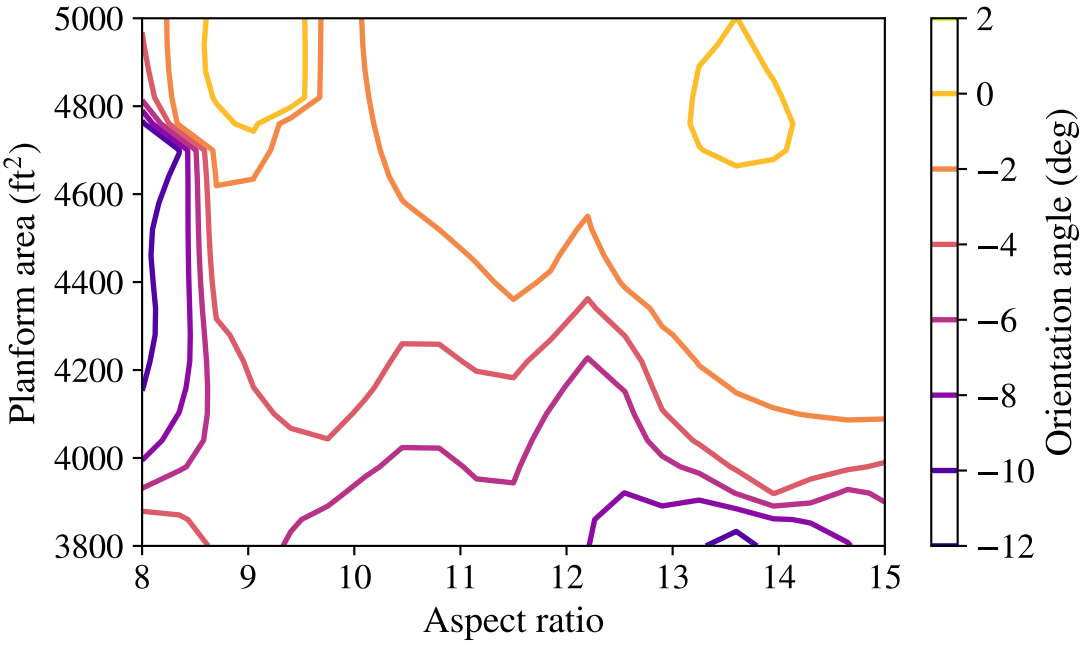

Figure 6.22: Optimal settings of lower orientation coefficient 3

the weight penalty from performing a local rather than global search is a quite small 0.2%. There are a couple of locations in the cusp region that also have a penalty, but these were only observed at individual points rather than in several neighboring points. The largest one corresponds to a weight penalty of 0.45%.

Aeroelasticity

Finally, the impact of including aeroelastic effects in the physics-based model was analyzed. The surrogate model's hyperparameters already indicate that not including aeroelasticity could lead to significant weight overestimation, but that could still be useful if the overestimation is consistent. Figure 6.24 show the relative difference between tow-steered and straight-tow designs, same as figure 6.19, but with no aeroelastic interaction (AEQR set to 0). Without aeroelasticity, the typical weight reduction is estimated to be 2.1% rather than 1.4%. Worse, the shape of the response over the planform space is different. Quantitatively,

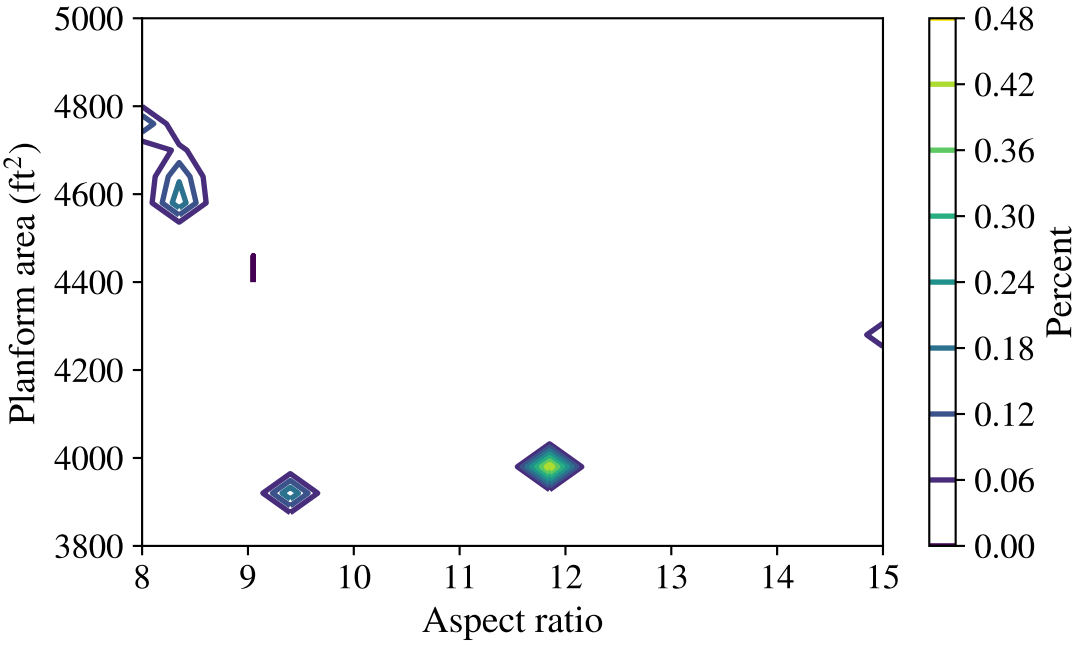

Figure 6.23: Percent difference between locally- and globally-optimized designs

the correlation coefficient between the percent change with and without aeroelasticity is 0.57.

Figure 6.25 further illustrates the difference by showing the slice of the planform space where the area is 4,160 ft^2. The aeroelastic and nonaeroelastic curves have similar gross downwards trends with increasing aspect ratio, but altogether are not that alike.

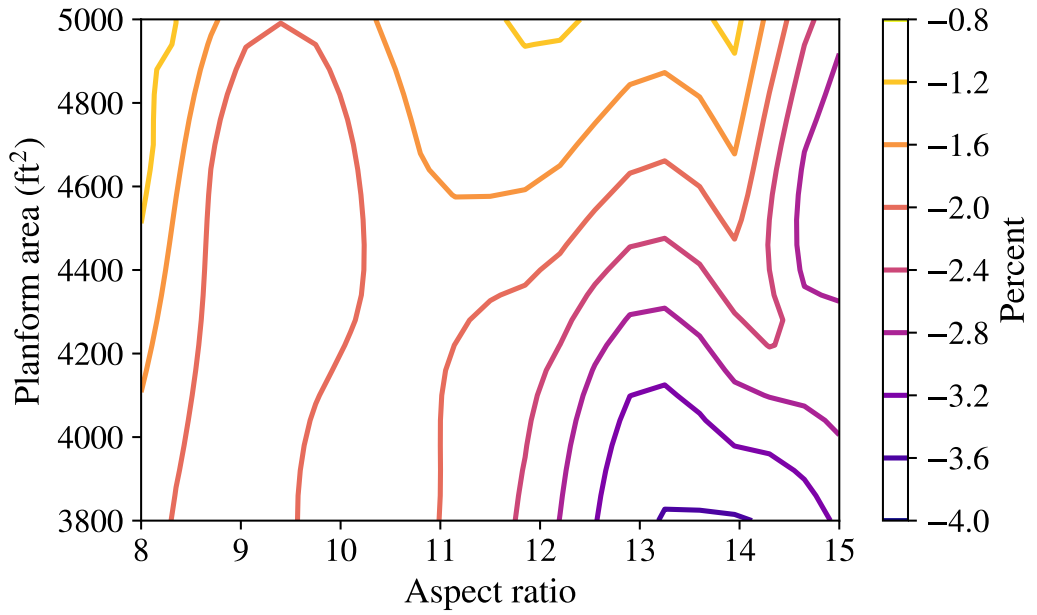

Figure 6.24: Contours of percent difference of tow-steered weight relative to straight-tow weight without aeroelasticity

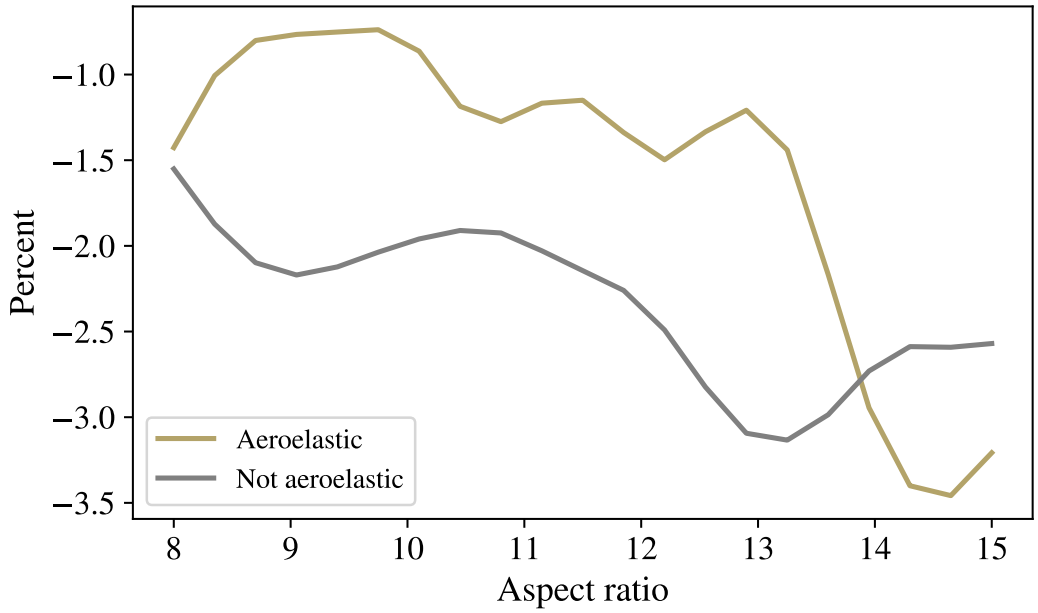

Figure 6.25: Sweep of percent difference in weight from tow steering with and without accounting for aeroelastic effects at a planform area of 4,160 ft^2

CHAPTER 7
CONCLUSION

A methodology for evaluating the weight reduction provided by tow-steered composites across a range of wing planforms has been developed and presented. The methodology produces a fast-running surrogate model that can be optimized within a system-level design process in order to support rapid iteration of the vehicle configuration with increased accuracy of the technology effect.

This methodology was developed after considering the results of two research focuses. The first was motivated by the high-dimensional parameterizations found in other work on tow-steered composite wingboxes, and sought to find a more efficient parameterization capable of obtaining the technology benefit with fewer design variables. This led to the development of an active subspace method for constrained optimization problems. An experiment was performed to determine whether this new method is able to achieve its intended goal, and a simple physics-based testbed was created in order to ensure that the experiment's results would be relevant. The results indicated that the method was working correctly but was ultimately unsuccessful, and as a result the methodology must compromise by using a lower-dimensional parameterization with less possible technology benefit in order to enable surrogate modeling.

The second research focus was motivated by the need to efficiently select data points for building the surrogate model in order to maximize the technology benefit while still producing an accurate surrogate. This led to the development of a new kind of Bayesian optimization method specifically intended for parameterized optimization problems. Two experiments were performed to test this new method. The first examined its performance on a variety of artificial test functions in order to determine how a function's characteristics affect the method's success. The second experiment applied the method to the existing

physics-based testbed to determine if it works for the intended purpose and to gain insight about how to formulate a tractable problem. These experiments showed the Bayesian parameterized optimization method to be successful, and therefore it was included in the thesis methodology.

After creating the methodology, an overarching hypothesis stating that it satisfies the research objective was made. Although exhaustively proving the overarching hypothesis is beyond the scope of this work, an example use case was presented to further substantiate it. This use case applied the methodology to a more realistic vehicle configuration, and demonstrated that the methodology was able to estimate the tow-steered composite benefit across the range of potential wing planforms.

In comparison to other tow-steered composite design approaches from literature, the methodology is limited to a coarser parameterization of the tow paths because it is based on surrogate modeling. As a result it might not be able to capture the full technology benefit. However, the methodology allows for the use of physics-based models that do not provide derivatives efficiently, enables global optimization of the tow paths, and can quickly yield designs across the planform design space. This complements those other methods nicely; for instance, the methodology could be used early in design while the vehicle configuration is more fluid to support system-level decision making and to gain an understanding of the design space, and later on a more richly parameterized physics-based model with efficient derivatives can be developed to take full advantage of tow steering. This second phase would benefit from the knowledge about different local minima gained in the first phase.

Bayesian parameterized optimization was critical for the success of a surrogate-based approach: in both experiment 3 and this use case, the initial data was collected using a space-filling design of experiments whose size was determined by a standard rule of thumb—in both applications the initial surrogate model had substantial error. The literature on surrogate modeling suggests that in such a situation, continuing to add data in a space-filling fashion would decrease the error at an impractically slow rate. Bayesian parameterized

optimization was able to identify important regions of the design space and focus data collection to those locations.

The use case also showcased RADE, a software library co-developed by the author for the purpose of building computational environments for structural weight estimation. The domain knowledge and automation capability embodied by such a software package is critical to being able to accurately assess new vehicle configurations and structural technologies earlier in the aircraft design process.

7.1 Findings

The results from experiment 1, experiment 3, and the example use case also shed some light on the behavior of tow-steered composite technology. Although tow steering can provide concrete benefits to the structure, the tow path variables have a noticeably smaller impact than the structural dimensions. This is clearly seen in figure 3.7, where the covariance associated with the tow orientations is much weaker than that associated with panel thicknesses. Additionally, the optimal tow paths tend to be aligned with the spars and do not take very extreme orientation angles. This makes sense because the internal load is primarily directed spanwise. This observation allowed the relatively wide design ranges that were initially used in experiment 3 to be narrowed, which made it easier to develop an accurate surrogate model. An important caveat is that if a more flexible parameterization were used, then the optimal tow paths might take on more extreme orientations in localized parts of the structure.

Another finding is that the strength of tow steering's effect varies with location along the wing span, with the greatest influence occurring just outboard of the wing root. Figure 3.7 also supports this point by showing that the tow orientation covariance is strongest in this region. Because of this pattern in the covariance, the first active subspace eigenvector in figure 3.10 can be seen to have a slight tow orientation component at the inboard wing. The importance of this area is also indicated by the sensitivity hyperparameters that the surrogate models learned from the data, seen in figures 4.21, 4.27 and 6.17. These figures all show

that, while tow steering has an effect throughout the wing excluding the tip, the section just outboard of the root has the strongest effect. Although the tow path parameterizations usually seen in the literature and used in this thesis consist of splines with evenly spaced knots, more useful design freedom per design variable could be obtained by skewing the knot distribution to favor this region.

The use case also showed that multiple local minima are possible in the tow-steered composite design space, though in most cases a gradient-based algorithm started from a default straight-tow design is likely to find the global minimum. Also, because the effect of tow steering is small, the difference between separate local minima is also small.

Finally, the results of the example use case can be used to infer how much each individual effect of tow steering contributes to its overall benefit to a CRM-like aircraft. First, coarse improvement of the load paths provides about a one to four percent weight reduction, depending on the planform. This can be concluded because the weight reduction benefit was observed both with and without aeroelasticity included in the physics-based model. Second, passive aeroelastic tailoring does not provide much benefit, but should be accounted for when quantifying the technology effect as a function of the planform. Comparing figures 6.19 and 6.24 shows that the technology benefit has similar size and overall trends regardless of whether aeroelasticity is modeled, but also that the specific planform-dependent benefit is different. Therefore, one should account for aeroelasticity when quantifying the technology benefit. However, if the benefit were approximated by a point estimate at a single planform, then the error associated with planform differences would probably dominate the error associated with excluding aeroelasticity from the physics-based model.

To complete the breakdown of tow steering's overall benefit, it should be noted that the technology benefit observed in the example use case is much less than other results for the CRM from literature, like [41] which reported a nine percent weight reduction. This suggests that in the use case the thesis methodology was only able to capture around half of the total technology benefit, though this conclusion is uncertain without directly comparing

the approaches on the same physics-based model. This difference might be due to the less flexible parameterization needed to keep the dimensionality small enough for surrogate modeling, and due to the tow orientation being set per skin panel rather than at the individual finite element level. This explanation is somewhat contradicted by figure 3.5, which shows that for the simple testbed, seven tow orientation parameters for each skin should be enough to capture the entire benefit from tow steering. However, that testbed has some significant simplifications, and the flexibility of the tow paths might become much more important as the realism of the physics-based model increases.

If the benefit from tow steering is largely dependent on both a flexible tow path parameterization and the details present in realistic structures, then the true benefit could be even larger than the values reported by studies that followed a high-dimensional gradient-based approach. Other work [40] has considered tow steering at the much smaller and more detailed scale of a single panel with a hole, and found that tow steering is able to redirect load away from the center of panels to better-supported features like edges or stiffeners. This improves resistance to buckling and reduces the need for pad-up structure around stress concentrations like cutouts and joints. This implies that structural details like cutouts, access panels, and joints need to be included in the physics-based model, and combined with a high dimensional tow path parameterization, in order to capture the full technology benefit. It also calls into question the homogenized panel stiffness approach, since it is not able to represent tow steering's ability to reroute load to individual stiffeners.

These details are usually not included in early-design models because that would greatly increase the amount of development effort. For other structural weight estimation projects, these details can be accounted for by calibration factors that modify the results of the necessarily simplified physics-based model. Unfortunately, this line of reasoning suggests that a calibration factor approach would not be adequate for tow steering. If the calibration factors in part account for stress concentrations not present in the model, and if tow steering allows stress concentrations to be dealt with more elegantly and efficiently, then the calibration

factors could be significantly different with and without tow steering. This would mean that accurately quantifying tow steering's full benefit is quite intensive, and would point to a need for improved capability to generate detailed structural models early in the vehicle design process.

7.2 Future Work

Active Subspaces

There are several ways that this thesis work could be extended. Although the active subspace method for constrained optimization problems from chapter 3 did not make its way into the methodology, it was able to learn several real features of the problem. Further attempts could be made to reduce the dimensionality, such as combining the handful of sensible active subspace directions with design basis directions obtained from another method. It is also possible that the active subspace method could be successful if applied to a different engineering problem involving constrained optimization.

Bayesian Parameterized Optimization

Bayesian parameterized optimization was found to be a successful method for efficiently constructing a surrogate model suitable for recovering optimal tow paths. However, it is not particularly specialized to the tow-steered composite design problem. Just as Bayesian optimization has been successfully applied to problems across many domains, it is likely that Bayesian parameterized optimization could be successfully applied to other low-dimensional parameterized optimization problems with expensive objective functions. It is therefore worthwhile to try to identify problems whose solution could be found more easily with the help of this method.

Like with tow-steered composites, such a problem can arise when a new technology introduces a set of design variables that are important for maximizing the technology's performance, but do not have a strong influence on other aspects of the system design

aside from the technology's performance. Other candidates are multidisciplinary design problems in which some of the variables associated with a discipline have only weak or no coupling effects with the other disciplines, making it possible to optimize them in separate subproblems. An example very similar to the one considered in this thesis is wing structural weight estimation while optimizing wingbox layout variables and allowing for different vehicle configuration possibilities. Wingbox layout optimization is especially suited to a surrogate-based approach because the rib spacing and the potential presence of a middle spar lead to discrete changes in the structural topology that are a challenge to address with gradient-based optimization. An approach similar to the thesis methodology could be useful for advanced or unconventional vehicle concepts, for which there might be less availability and development of physics-based models. Bayesian parameterized optimization combined with fully stressed design would enable structural optimization with a physics-based model that does not provide derivatives, while also considering different vehicle configuration possibilities.

One of the difficulties of Bayesian optimization is estimation of the hyperparameters, as seen in chapter 4. When point estimates of the hyperparameters are used the acquisition function cannot account for hyperparameter uncertainty. A hierarchical Bayesian modeling approach, in which there is also a posterior distribution for the hyperparameters, could be more robust although the surrogate would also be more expensive to train.

It would be interesting to try variations on the integrated knowledge gradient acquisition function. For instance, the alternate form shown in equation (7.1b) would better favor points that provide a lot of surrogate improvement even if that improvement only applies to a small

region of parameter space.

$$\mathbb{IKG} = \int \mathbb{KG}\, \mathrm{d}\boldsymbol{t}' \tag{7.1a}$$

$$\mathbb{IKG}_{\mathrm{alt}} = \ln\left(\int [\exp(\mathbb{KG}) - 1]\, \mathrm{d}\boldsymbol{t}'\right) \tag{7.1b}$$

It also might be possible to devise an acquisition function that is more specialized for technology impact estimation. The integrated knowledge gradient reflects a goal of finding the optimal designs, but the true goal is to estimate the improvement in the optimal designs enabled by adding the technology. A more focused acquisition function might be able to reduce the amount of data required to obtain this answer. Rather than integrating the knowledge gradient, which measures the expected difference between the optimum of the current surrogate's mean prediction and the optimum of the next-stage surrogate's mean prediction, the integrand could be based on the expected difference between the technology benefit indicated by the current surrogate's mean prediction and the technology benefit indicated by the next-stage surrogate's mean prediction.

Thesis Methodology

An important result from the example use case was that the thesis methodology obtained a lower technology benefit than a competing approach from literature. A potential cause is the less flexible parameterization that the methodology must use to keep the dimensionality manageable, and which might be limiting opportunities to improve local resistance to load and structural failure.

One of the best ways to enhance this thesis work would be to definitively determine the cause of this discrepancy. This could be accomplished by directly comparing the thesis methodology to a conventional high-dimensional gradient-based optimization approach. However, the physics-based model would need to provide gradients in order to use the

conventional approach. The simple testbed from experiments 1 and 3 had this capability and facilitated a direct comparison, but its other simplifications could have distorted the estimated technology benefit. Therefore, a physics-based model of a realistic wing with gradient capabilities would be needed to further study the discrepancy.

If the less flexible parameterization is responsible for the lower observed technology benefit in the example use case, then there are a couple of ways to try to improve this aspect of the methodology. First, the portion of the technology benefit that the methodology cannot capture is based on rapidly varying tow orientations to obtain localized improvement. It might be reasonable to assume that localized improvement has relatively uniform effects over the structure, depending almost entirely on structural weight and not on planform shape. If that were true then the missing portion of the technology benefit could be accounted for with an additional calibration factor for local tow steering, though research would be needed to determine its value.

Another possible solution would be to keep the parameterization small enough to permit surrogate modeling, but redefine it so the tow path design is based on a mix of local and global considerations. Since most of tow steering's weight reduction benefit is due to better local resistance to load and failure, the tow paths could initially be aligned with the largest principal stresses of an important load case (e.g., a 2.5g pull-up, and possibly 1g push-down for the lower skin). This design should capture a large portion of the technology benefit attributed to local improvements, and since it is obtained from analysis results rather than from design variables, it does not need a high-dimensional parameterization. Research might be needed to determine the best way to initialize the structural properties and to stabilize the principal stress directions between iterations as the design is revised. It might also be helpful to average the principal stress directions over small regions of the structure. In order to also capture larger-scale load redirection and passive aeroelastic tailoring, this stress-aligned orientation field could then be perturbed by another orientation field parameterized by a low-dimensional spline, like in the thesis methodology. If successful, this approach would

allow for flexible tow paths that capture the large benefit from better local fiber alignment, would also capture the benefits from improved load paths and passive aeroelastic tailoring, and would use a low-dimensional parameterization that permits surrogate modeling.

Finally, the methodology would benefit from a systematic approach to choosing the tow path parameterization and design ranges. The experiments and use case in this thesis used B-splines with evenly spaced knots and took a simple approach to setting design ranges. However, there is evidence that this approach was inefficient, like the fact that the wingtip tow path coefficients did not have much effect on the weight. Essentially, some of the limited design freedom was wasted by overallocating it to the wingtip. A better approach would provide a more intelligent way to select the locations of the spline knots and the design ranges for the spline coefficients. This would enable the methodology to capture more of tow steering's potential benefit with a limited number of variables.

REFERENCES

[1] "Commercial market outlook 2020–2039," Boeing, Tech. Rep., Oct. 2020.

[2] F. Collier, R. Thomas, C. Burley, C. Nickol, C.-M. Lee, and M. Tong, "Environmentally Responsible Aviation—Real solutions for environmental challenges facing aviation," in *27th International Congress of the Aeronautical Sciences*, (Nice, France), Sep. 2010.

[3] R. Del Rosario, "NASA Aeronautics Advanced Air Transport Technology (AATT)," in *Applied Modeling & Simulation (AMS) Seminar Series*, (Mountain View, CA), NASA, May 2015.

[4] *Fact sheet—Continuous Lower Energy, Emissions, and Noise (CLEEN) program*, Web Article, 2019.

[5] J. Schutte, H. Jimenez, and D. Mavris, "Technology assessment of NASA Environmentally Responsible Aviation advanced vehicle concepts," in *49th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, (Orlando, FL), AIAA, Jan. 2011.

[6] NASA Aeronautics Research Mission Directorate, "NASA Aeronautics—Strategic Implementation Plan—2019 Update," NASA, Washington, DC, NASA Publication NP-2017-01-2352-HQ, 2019.

[7] D. P. Raymer, "Overview of the design process," in *Aircraft Design: A Conceptual Approach*, 4th ed., ser. AIAA Education Series. Reston, VA: AIAA, 2006, ch. 2, pp. 3–14.

[8] G. L. Giles, "Design oriented structural analysis," NASA, Hampton, VA, Technical Memorandum NASA–TM–109124, Jun. 1994.

[9] L. A. McCullers, "Aircraft configuration optimization including optimized flight profiles," in *Recent Experiences in Multidisciplinary Analysis and Optimization*, (Hampton, VA), J. Sobieski, Ed., Washington, DC: NASA, Apr. 1984, pp. 395–412.

[10] M. R. Kirby and D. N. Mavris, "The Environmental Design Space," in *26th International Congress of the Aeronautical Sciences*, (Anchorage, AK), Sep. 2008.

[11] ——, "A method for technology selection based on benefit, available schedule and budget resources," in *World Aviation Conference*, (San Diego, CA), Washington, DC: AIAA, Oct. 2000.

[12] J. S. Schutte and D. N. Mavris, "Evaluation of N+2 technologies and advanced vehicle concepts," in *53rd AIAA Aerospace Sciences Meeting*, (Kissimmee, FL), AIAA, Jan. 2015.

[13] M. R. Kirby, "A methodology for technology identification, evaluation, and selection in conceptual and preliminary aircraft design," PhD thesis, Georgia Institute of Technology, Atlanta, GA, Mar. 2001.

[14] O. Dababneh and T. Kipouros, "A review of aircraft wing mass estimation methods," *Aerospace Science and Technology*, vol. 72, pp. 256–266, Jan. 2018.

[15] A. Elham, G. La Rocca, and M. J. L. van Tooren, "Development and implementation of an advanced, design-sensitive method for wing weight estimation," *Aerospace Science and Technology*, vol. 29, no. 1, pp. 100–113, Aug. 2013.

[16] M. D. Ardema, M. C. Chambers, A. P. Patron, A. S. Hahn, H. Miura, and M. D. Moore, "Analytical fuselage and wing weight estimation of transport aircraft," NASA, Washington, DC, Technical Memorandum NASA–TM–110392, May 1996.

[17] N. Murphy, "Analytical wing weight prediction/estimation using computer based design techniques," PhD thesis, Cranfield Institute of Technology, Cranfield, England, 1987.

[18] M. C.-Y. Niu, "Weight control and balance," in *Airframe Structural Design—Practical Design Information and Data on Aircraft Structures*, 2nd ed. North Point, Hong Kong: Hong Kong Conmilit Press, 1999, ch. 16.

[19] D. P. Raymer, "Weights," in *Aircraft Design: A Conceptual Approach*, 4th ed., ser. AIAA Education Series. Reston, VA: AIAA, 2006, ch. 15, pp. 451–466.

[20] W. R. Micks, "Structural weight analysis—Wing weight equations," The Rand Corporation, Santa Monica, CA, Tech. Rep. R-198, Dec. 1950.

[21] D. P. Wells, B. L. Horvath, and L. A. McCullers, "The Flight Optimization System weights estimation method," NASA, Washington, DC, Technical Memorandum TM–2017–219627, Jun. 2017.

[22] E. Torenbeek, "Prediction of wing group weight for preliminary design," *Aircraft Engineering and Aerospace Technology*, vol. 43, no. 7, pp. 16–21, Jul. 1971.

[23] N.-H. Kim and B. V. Sankar, *Introduction to Finite Element Analysis and Design*. New York, NY: John Wiley & Sons, Inc, 2009.

[24] G. L. Giles, "Equivalent plate modeling for conceptual design of aircraft wing structures," in *1st AIAA Aircraft Engineering, Technology and Operations Congress*, (Los Angeles, CA), Washington, DC: AIAA, Sep. 1995.

[25] A. Elham and M. J. L. van Tooren, "Tool for preliminary structural sizing, weight estimation, and aeroelastic optimization of lifting surfaces," *Journal of Aerospace Engineering*, vol. 230, no. 2, pp. 280–295, 2016.

[26] G. Bindolino, G. Ghiringhelli, S. Ricci, and M. Terraneo, "Multilevel structural optimization for preliminary wing-box weight estimation," *Journal of Aircraft*, vol. 47, no. 2, pp. 475–489, Mar. 2010.

[27] S. Venkataraman and R. T. Haftka, "Structural optimization complexity: What has Moore's law done for us?" *Structural and Multidisciplinary Optimization*, vol. 28, no. 6, pp. 375–387, Sep. 2004.

[28] J. A. Corman, "A methodology for structural technology performance characterization to enable reduction of structural uncertainty," PhD thesis, Georgia Institute of Technology, Atlanta, GA, May 2017.

[29] *Part 25—Airworthiness standards: Transport category airplanes*, Electronic Code of Federal Regulations, Dec. 2019.

[30] A. Velicki, "Damage arresting composites for shaped vehicles—Phase 1 final report," NASA, Washington, DC, Contractor Report NASA/CR–2009–215932, Sep. 2009.

[31] C. S. Collier, "Thermoelastic formulation of stiffened, unsymmetric composite panels for finite element analysis of high speed aircraft," in *35th Structures, Dynamics, and Materials Conference*, (Hilton Head, SC), Washington, DC: AIAA, Apr. 1994.

[32] M. Sensmeier, B. Stewart, and J. Samareh, "Rapid generation and assessment of aircraft structural topologies for multidisciplinary optimization and weight estimation," in *47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, (Newport, RI), AIAA, May 2006.

[33] R. Del Rosario, S. Anders, N. Madavan, G. Martin, S. Helland, and S. Viken, "Advanced Air Vehicles Program (AAVP)—Advanced Air Transport Technology project (AATT)," (Cleveland, OH), NASA, Mar. 2015.

[34] O. Stodieck, J. E. Cooper, P. M. Weaver, and P. Kealy, "Aeroelastic tailoring of a representative wing box using tow-steered composites," *AIAA Journal*, vol. 55, no. 4, pp. 1425–1439, Apr. 2017.

[35] J. Levine, "Experimental wing verified during loads testing," *NASA*, Nov. 2018.

[36] K. C. Wu, B. F. Tatting, B. H. Smith, R. S. Stevens, G. P. Occhipinti, J. B. Swift, D. C. Achary, and R. P. Thornburgh, "Design and manufacturing of tow-steered composite shells using fiber placement," in *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, (Palm Springs, CA), AIAA, May 2009.

[37] T. R. Brooks and J. R. R. A. Martins, "On manufacturing constraints for tow-steered composite design optimization," *Composite Structures*, vol. 204, pp. 548–559, Nov. 2018.

[38] B. C. Kim, K. Potter, and P. M. Weaver, "Continuous tow shearing for manufacturing variable angle tow composites," *Composites Part A: Applied Science and Manufacturing*, vol. 43, no. 8, pp. 1347–1356, Aug. 2012.

[39] B. K. Stanford and C. V. Jutte, "Aeroelastic tailoring via tow steered composites," NASA, Washington, DC, Technical Memorandum NASA/TM–2014–218517, Sep. 2014.

[40] C. S. Lopes, Z. Gürdal, and P. P. Camanho, "Tailoring for strength of composite steered-fibre panels with cutouts," *Composites Part A: Applied Science and Manufacturing*, vol. 41, no. 12, pp. 1760–1767, Dec. 2010.

[41] T. R. Brooks, J. R. R. A. Martins, and G. J. Kennedy, "High-fidelity multipoint aerostructural optimization of a high aspect ratio tow-steered composite wing," in *58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, (Grapevine, TX), AIAA, Jan. 2017.

[42] M. Rouse, D. C. Jegley, D. M. McGowan, H. G. Bush, and W. A. Waters, "Utilization of the building-block approach in structural mechanics research," in *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, (Austin, TX), AIAA, Apr. 2005.

[43] "Guidelines for property testing of composites," in *Department of Defense Handbook—Composite Materials Handbook—Polymer Matrix Composites Guidelines for Characterization of Structural Materials MIL-HDBK-17-1F*. DOD, 2002, vol. 1, ch. 2.

[44] T. R. Brooks, J. T. Hwang, G. J. Kennedy, and J. R. R. A. Martins, "High-fidelity structural optimization of a tow-steered composite wing," in *11th World Conference on Structural and Multidisciplinary Optimization*, (Syndey, Australia), Jun. 2015.

[45] R. Razani, "Behavior of fully stressed design of structures and its relationship to minimum-weight design," *AIAA Journal*, vol. 3, no. 12, pp. 2262–2268, 1965.

[46] B. Liu, R. T. Haftka, and L. T. Watson, "Global-local structural optimization using response surfaces of local optimization margins," *Structural and Multidisciplinary Optimization*, vol. 27, no. 5, pp. 352–359, Jun. 2004.

[47] R. M. Pickett Jr., M. F. Rubinstein, and R. B. Nelson, "Automated structural synthesis using a reduced number of design coordinates," in *AIAA/ASME/SAE 14th Structures, Structural Dynamics, and Materials Conference*, (Williamsburg, VA), New York, NY: AIAA, Mar. 1973.

[48] P. G. Constantine, *Active Subspaces—Emerging Ideas for Dimension Reduction in Parameter Studies*, ser. SIAM Spotlights. Philadelphia, PA: SIAM, 2015.

[49] P. G. Constantine and D. F. Gleich, "Computing active subspaces with Monte Carlo," *arXiv:1408.0545v2 [math.NA]*, Jul. 2015.

[50] T. Lukaczyk, F. Palacios, J. J. Alonso, and P. G. Constantine, "Active subspaces for shape optimization," in *10th AIAA Multidisciplinary Design Optimization Conference*, (National Harbor, MD), AIAA, Jan. 2014.

[51] S. H. Berguin, D. Rancourt, and D. N. Mavris, "Method to facilitate high-dimensional design space exploration using computationally expensive analyses," *AIAA Journal*, vol. 53, no. 12, pp. 3752–3765, Dec. 2015.

[52] J. Nocedal and S. J. Wright, "Penalty and augmented Lagrangian methods," in *Numerical Optimization*, T. V. Mikosch, S. I. Resnick, and S. M. Robinson, Eds., 2nd ed., ser. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer, 2006, ch. 17, pp. 497–528.

[53] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming—Sequential Unconstrained Minimization Techniques*, ser. Classics in Applied Mathematics. Philadelphia, PA: SIAM, 1990.

[54] P. G. Constantine, "Optimization," in *Active Subspaces—Emerging Ideas for Dimension Reduction in Parameter Studies*, ser. SIAM Spotlights. Philadelphia, PA: SIAM, 2015, ch. 4.4, pp. 62–65.

[55] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Belmont, MA: Athena Scientific, 1982.

[56] D. Goldfarb and A. Idnani, "A numerically stable dual method for solving strictly convex quadratic programs," *Mathematical Programming*, vol. 27, pp. 1–33, Sep. 1983.

[57] R. H. Byrd, R. B. Schnabel, and G. A. Shultz, "A trust region algorithm for nonlinearly constrained optimization," *SIAM Journal on Numerical Analysis*, vol. 24, no. 5, pp. 1152–1170, Oct. 1987.

[58] Y. A. Çengel and J. M. Cimbala, "Differential analysis of fluid flow," in *Fluid Mechanics—Fundamentals and Applications*, 2nd ed. New York, NY: McGraw-Hill, 2010, ch. 9, pp. 419–490.

[59] O. Schrenk, "A simple approximation method for obtaining the spanwise lift distribution," NACA, Washington, DC, Technical Memorandum 948, Aug. 1940.

[60] S. G. Hedman, "Vortex lattice method for calculation of quasi steady state loadings on thin elastic wings in subsonic flow," The Aeronautical Research Institute of Sweden, Stockholm, Sweden, Report 105, 1966.

[61] J. J. Bertin and R. M. Cummings, "Vortex lattice method," in *Aerodynamics for Engineers*, 5th ed. Upper Saddle River, NJ: Pearson Prentice-Hall, 2009, ch. 7.5.

[62] O. Bauchau and J. Craig, "Basic equations of linear elasticity," in *Structural Analysis—With Applications to Aerospace Structures*, ser. Solid Mechanics and its Applications. New York, NY: Springer, 2009, ch. 1.

[63] A. T. Nettels, "Basic mechanics of laminated composite plates," NASA, Washington, DC, Reference Publication NASA–RP–1351, Oct. 1994.

[64] Z. Gürdal, R. T. Haftka, and P. Hajela, "Mechaniccs of laminated composite materials," in *Design and Optimization of Laminated Composite Materials*. New York, NY: John Wiley & Sons, 1999, ch. 2.

[65] R. L. Harder and R. N. Desmarais, "Interpolation using surface splines," *Journal of Aircraft*, vol. 9, no. 2, pp. 189–191, Feb. 1972.

[66] S. W. Tsai and E. M. Wu, "A general theory of strength for anisotropic materials," *Journal of Composite Materials*, vol. 5, no. 1, pp. 58–80, Jan. 1971.

[67] A. W. Leissa, "Buckling of laminated composite plates and shell panels," Flight Dynamics Laboratory, Wright-Patterson Air Force Base, OH, Final Report AFWAL–TR–85–3069, Jun. 1985.

[68] M. C.-Y. Niu, "Buckling and stability of structures," in *Airframe Structural Design—Practical Design Information and Data on Aircraft Structures*, 2nd ed. North Point, Hong Kong: Hong Kong Conmilit Press, 1999, ch. 5.

[69] "Compressive buckling and crippling," in *Department of Defense Handbook—Composite Materials Handbook—Polymer Matrix Composites Materials Usage, Design, and Analysis MIL-HDBK-17-3F*. DOD, 2002, vol. 3, ch. 5.7.

[70] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed., T. V. Mikosch, S. I. Resnick, and S. M. Robinson, Eds., ser. Springer Series in Operations Research and Financial Engineering. New York, NY: Springer, 2006.

[71] J. Iott, R. T. Haftka, and H. M. Adelman, "Selecting step sizes in sensitivity analysis by finite differences," NASA, Hampton, VA, Technical Memorandum NASA-TM-86382, Aug. 1985.

[72] R. Mathur, "An analytic approach to computing step sizes for finite-difference derivatives," PhD thesis, University of Texas at Austin, Austin, TX, May 2012.

[73] A. Griewank and A. Walther, *Evaluating Derivatives—Principles and Techniques of Algorithmic Differentiation*, 2nd ed. Philadelphia, PA: SIAM, 2008.

[74] A. B. Lambe, G. J. Kennedy, and J. R. R. A. Martins, "An evaluation of constraint aggregation strategies for wing box mass minimization," *Structural and Multidisciplinary Optimization*, vol. 55, no. 1, pp. 257–277, Jan. 2017.

[75] Python Software Foundation, *Python*, Software.

[76] MSC Software, *MSC Nastran*, Software, version 2017.0, 2017.

[77] S. Doyle, *pyNastran*, Software.

[78] J. C. Vassberg, M. A. DeHaan, S. M. Rivers, and R. A. Wahls, "Development of a Common Research Model for applied CFD validation studies," in *26th AIAA Applied Aerodynamics Conference*, (Honolulu, HI), AIAA, Aug. 2008.

[79] G. Kenway, G. Kennedy, and J. Martins, "Aerostructural optimization of the Common Research Model configuration," in *15th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, (Atlanta, GA), AIAA, Jun. 2014.

[80] Open Cascade, *Open CASCADE Technology*, Software.

[81] OPEN CASCADE, *SMESH*, Software.

[82] T. Laughlin, *pyOCCT*, Software, version 7.4.0, 2020.

[83] MSC Software, *Design Sensitivity and Optimization User's Guide*. Newport Beach, CA: MSC Software, 2016.

[84] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, l. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[85] D. Kraft, "A software package for sequential quadratic programming," DFVLR, Koeln, Germany, Tech. Rep. DFVLR-FB–88-28, 1988.

[86] G. N. Vanderplaats, "Structural design," in *Multidiscipline Design Optimization*, 1st ed. Monterey, CA: Vanderplaats Research & Development, Inc., 2007, ch. 9-4, pp. 315–339.

[87] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook, *Response Surface Methodology—Process and Product Optimization Using Designed Experiments*, 3rd ed., ser. Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons, Inc., 2009.

[88] G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters—An Introduction to Design, Data Analysis, and Model Building*, ser. Wiley Series in Probability and Mathematical Statisics. New York, NY: John Wiley & Sons, Inc., 1978.

[89] T. J. Santner, B. J. Williams, and W. I. Notz, "Space-filling designs for computer experiments," in *The Design and Analysis of Computer Experiments*, 2nd ed., ser. Springer Series in Statistics. New York, NY: Springer, 2018, ch. 5.

[90] W. J. Welch, R. J. Buck, J. Sacks, H. P. Wynn, M. D. Morris, and M. Schonlau, "Response to James M. Lucas," *Technometrics*, vol. 38, no. 2, pp. 199–203, May 1996.

[91] C. E. Rasmussen and K. I. Williams, *Gaussian Processes for Machine Learning*, ser. Adaptive Computation and Machine Learning. Cambridge, MA: MIT Press, 2006.

[92] F. A. C. Viana, G. Venter, and V. Balabanov, "An algorithm for fast optimal Latin hypercube design of experiments," *International Journal for Numerical Methods in Engineering*, vol. 82, pp. 135–156, 2010.

[93] J. L. Loeppky, J. Sacks, and W. J. Welch, "Choosing the sample size of a computer experiment: A practical guide," *Technometrics*, vol. 51, no. 4, pp. 366–376, Nov. 2009.

[94] L. M. Benveniste and J. A. Scheinkman, "On the differentiability of the value function in dynamic models of economics," *Econometrica*, vol. 47, no. 3, 727–732, May 1979.

[95] H. Liu, Y.-S. Ong, and J. Cai, "A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design," *Structural and Multidisciplinary Optimization*, vol. 57, no. 1, pp. 393–416, 2018.

[96] C. Q. Lam and W. I. Notz, "Sequential adaptive designs in computer experiments for response surface model fit," *Statistics and Applications*, vol. 6, no. 1&2, pp. 207–233, 2008.

[97] A. I. J. Forrester and A. J. Keane, "Recent advances in surrogate-based optimization," *Progress in Aerospace Sciences*, vol. 45, pp. 50–79, Jan. 2009.

[98] R. T. Haftka, D. Villanueva, and A. Chaudhuri, "Parallel surrogate-assisted global optimization with expensive functions—a survey," *Structural and Multidisciplinary Optimization*, vol. 54, no. 1, pp. 3–13, 2016.

[99] E. Iuliano, "Efficient design optimization assisted by sequential surrogate models," *International Journal of Aerospace Engineering*, vol. 2019, pp. 1–34, May 2019.

[100] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, pp. 455–492, 1998.

[101] D. Huang, T. T. Allen, W. I. Notz, and N. Zeng, "Global optimization of stochastic black-box systems via sequential Kriging meta-models," *Journal of Global Optimization*, vol. 34, no. 3, pp. 441–466, 2006.

[102] P. Ranjan, D. Bingham, and G. Michailidis, "Sequential experiment design for contour estimation from complex computer codes," *Technometrics*, vol. 50, no. 4, pp. 527–541, Nov. 2008.

[103] C. E. Rasmussen and Z. Ghahramani, "Bayesian Monte Carlo," in *15th International Conference on Neural Information Processing Systems*, (Vancouver, BC), S. Becker, S. Thrun, and K. Obermayer, Eds., Cambridge, MA: MIT Press, Dec. 2002, pp. 505–512.

[104] B. J. Bichon, J. M. McFarland, and S. Mahadevan, "Efficient surrogate models for reliability analysis of systems with multiple failure modes," *Reliability Engineering & System Safety*, vol. 96, no. 10, pp. 1386–1395, 2011.

[105] B. J. Williams, T. J. Santner, and W. I. Notz, "Sequential design of computer experiments to minimize integrated response functions," *Statistica Sinica*, vol. 10, pp. 1133–1152, 2000.

[106] K. Swersky, J. Snoek, and R. P. Adams, "Multi-task Bayesian optimization," in *Advances in Neural Information Processing Systems 26*, 2013.

[107] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *25th International Conference on Neural Information Processing Systems*, (Lake Tahoe, NV), vol. 2, Red Hook, NY: Curran Associates Inc., Dec. 2012, pp. 2951–2959.

[108] Y. Wang, K. G. Reyes, K. A. Brown, C. A. Mirkin, and W. B. Powell, "Nested-batch-mode learning and stochastic optimization with an application to sequential multistage testing in materials science," *SIAM Journal on Scientific Computing*, vol. 37, no. 3, B361–B381, 2015.

[109] R. Lam, M. Poloczek, P. I. Frazier, and K. E. Willcox, "Advances in Bayesian optimization with applications in aerospace engineering," in *2018 AIAA Non-Deterministic Approaches Conference*, (Kissimmee, FL), AIAA, Jan. 2018.

[110] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed., ser. Texts in Statistical Science. Boca Raton, FL: Chapman & Hall/CRC, 2013.

[111] M. D. Homan and A. Gelman, "The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1593–1623, Apr. 2014.

[112] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*, 2nd ed., ser. Springer Series in Statistics. New York, NY: Springer, 2018.

[113] C. E. Rasmussen and C. K. I. Williams, "Covariance functions," in *Gaussian Processes for Machine Learning*, ser. Adaptive Computation and Machine Learning. Cambridge, MA: MIT Press, 2006, ch. 4.

[114] ——, "Model selection and adaptation of hyperparameters," in *Gaussian Processes for Machine Learning*, ser. Adaptive Computation and Machine Learning. Cambridge, MA: MIT Press, 2006, ch. 5.

[115] A. Mockus, J. Mockus, and L. Mockus, "Bayesian approach adapting stochastic and heuristic methods of global and discrete optimization," *Informatica*, vol. 5, no. 1-2, pp. 123–166, 1994.

[116] D. D. Cox and S. John, "SDO: A statistical method for global optimization," in *Multidisciplinary Design Optimization—State of the Art—Proceedings of the ICASE/NASA Langley Workshop on Multidisciplinary Design Optimization*, N. M. Alexandrov and M. Y. Hussaini, Eds. Philadelphia, PA: SIAM, 1997, pp. 315–329.

[117] P. Frazier, W. Powell, and S. Dayanik, "The knowledge-gradient policy for correlated normal beliefs," *INFORMS Journal on Computing*, vol. 21, no. 4, pp. 599–613, 2009.

[118] W. Scott, P. Frazier, and W. Powell, "The correlated knowledge gradient for simulation optimization of continuous parameters using Gaussian process regression," *SIAM Journal on Optimization*, vol. 21, no. 3, pp. 996–1026, Sep. 2011.

[119] J. M. Hernández-Lobato, M. W. Hoffman, and Z. Ghahramani, "Predictive entropy search for efficient global optimization of black-box functions," in *27th International Conference on Neural Information Processing Systems*, (Montreal, Canada), vol. 1, Dec. 2014, pp. 918–926.

[120] J. Wu and P. I. Frazier, "The parallel knowledge gradient method for batch Bayesian optimization," in *30th Conference on Neural Information Processing Systems*, (Barcelona, Spain), NIPS, Dec. 2016.

[121] A. Törn and A. Žilinskas, "Testing and applications," in *Global Optimization*, ser. Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 1989, ch. 8.

[122] J. M. Hammersley and K. W. Morton, "A new Monte Carlo technique: Antithetic variates," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 52, no. 3, pp. 449–475, Jul. 1956.

[123] V. Beiranvand, W. Hare, and Y. Lucet, "Best practices for comparing optimization algorithms," *Optimization and Engineering*, vol. 18, no. 4, pp. 815–848, Dec. 2017.

[124] R. Sala, N. Baldanzini, and M. Pierini, "Representative surrogate problems as test functions for expensive simulators in multidisciplinary design optimization of vehicle structures," *Structural and Multidisciplinary Optimization*, vol. 54, no. 3, pp. 449–468, Sep. 2016.

[125] C.-K. Ng and D. Li, "Test problem generator for unconstrained global optimization," *Computers & Operations Research*, vol. 51, pp. 338–349, Nov. 2014.

[126] B. Addis and M. Locatelli, "A new class of test functions for global optimization," *Journal of Global Optimization*, vol. 38, no. 3, pp. 479–501, Jul. 2007.

[127] M. Gaviano, D. E. Kvasov, D. Lera, and Y. D. Sergeyev, "Algorithm 829: Software for generation of classes of test functions with known local and global minima for

global optimization," *ACM Transactions on Mathematical Software*, vol. 29, no. 4, pp. 469–480, Dec. 2003.

[128] R. Bellman, "The curse of dimensionality," in *Adaptive Control Processes : A Guided Tour*, ser. Princeton Legacy Library. Princeton, NJ: Princeton University Press, 1961, ch. 5.16.

[129] D. J. Wales and J. P. Doye, "Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 100 atoms," *Journal of Physical Chemistry A*, vol. 101, no. 28, pp. 5111–5116, Jul. 1997.

[130] M. Shewry and H. Wynn, "Maximum entropy sampling," *Journal of Applied Statistics*, vol. 14, no. 2, pp. 165–170, 1987.

[131] "AS4 12k/3502 unitape," in *Department of Defense Handbook—Composite Materials Handbook—Polymer Matrix Composites Materials Properties MIL-HDBK-17-2E*. DOD, 1999, vol. 2, ch. 4.2.8.

[132] Collier Research Corporation, *HyperSizer*, Software, version 7.3.50, 2017.

[133] J. A. Corman, N. R. Weston, C. I. Friedland, D. N. Mavris, and T. W. Laughlin, "Rapid Airframe Design Environment (RADE): A parametric, modular, and multidisciplinary framework for conceptual phase airframe design," in *2018 AIAA Modeling and Simulation Technologies Conference*, (Kissimmee, FL), AIAA, Jan. 2018.

[134] J. Gloudemans, R. McDonald, M. Moore, A. Hahn, B. Fredericks, and A. Gary, *OpenVSP*, Software, version 3.5.2, 2016.

[135] T. Laughlin, *Airframe Finite Element Modeler*, Software, version 0.2.0, Laughlin Research, 2018.

[136] Netgen/NGSolve, *Netgen*, Software, version 6.2, 2017.

[137] M. Drela and H. Youngren, *AVL*, Software, version 3.36, 2017.

[138] A. Cox, *nastran_utils*, Software, version 0.6, 2018.

[139] *HyperSizer methods and equations—Thermoelastic stiffness formulation*, Collier Research Corporation, Aug. 2012.

[140] *HyperSizer methods and equations—FEA loads*, Collier Research Corporation, May 2013.

[141] *HyperSizer methods and equations—Plate and shell buckling*, Collier Research Corporation, Nov. 2010.

[142] *HyperSizer methods and equations—Panel buckling*, Collier Research Corporation, Sep. 2012.

[143] *HyperSizer methods and equations—Local buckling*, Collier Research Corporation, Mar. 2013.

[144] *HyperSizer methods and equations—Crippling of panels and beams*, Collier Research Corporation, May 2012.

[145] *HyperSizer methods and equations—Composite strength, ply-based*, Collier Research Corporation, Apr. 2015.

[146] B. K. Stanford, C. V. Jutte, and C. D. Wieseman, "Trim and structural optimization of subsonic transport wings using nonconventional aeroelastic tailoring," *AIAA Journal*, vol. 54, no. 1, pp. 293–309, Jan. 2016.

VITA

Coleby Friedland was born in Land O' Lakes, Florida in 1990. He took an interest in aircraft starting at an early age, so after graduating from Land O' Lakes High School in 2008 he enrolled at the University of Florida to pursue a Bachelor of Science in Aerospace Engineering. After earning that degree in 2012, he continued his studies by joining the Aerospace Systems Design Lab at the Georgia Institute of Technology. During his time in the lab, Coleby worked on projects sponsored by a variety of organizations including Boeing, Airbus, and NASA—most of which were focused on increasing structural knowledge early in aircraft design. Coleby was awarded a Master of Science in Aerospace Engineering in 2014 and a Doctor of Philosophy in 2021.